DOCUMENT RESUME

ED 143 675 95 TM 006 378

AUTHOR	Fennessey, James
TITLE	Blending Evidence, Technique, and Judgment in Educational Research Inference. Final Report.
INSTITUTION	Johns Hopkins Univ., Baltimore, Md. Dept. of Social Relations.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
BUREAU NO	NIE-74-0071
PUB DATE	Nov 76
NOTE	298p.; Some of the tables may be marginally legible
EDRS PRICE	MF-$0.83 HC-$15.39 Plus Postage.
DESCRI[illegible]ORS	*Bayesian Statistics; *Data Analysis; Disadvantaged Youth; *Educational Research; *Evaluation Methods; Learning; Mathematical Models; Research Design; Research Methodology

ABSTRACT

This final report of a National Institute of Education project explores Bayesian statistical analysis as a paradigm for educational impact studies, particularly studies on the education of the disadvantaged. The position of the report is that much of what is wrong with educational research can be attributed to the use of an inappropriate model for making inferences. The author presents and discusses advantages of Bayesian inference over the "sampling-theory" framework of inference. Particular attention is paid to the demonstration of advantages of the Bayesian paradigm as a basis for representing knowledge when the data are "weak." Computer programs are provided to facilitate the application of Bayesian analysis to the sort of data most frequently encountered in educational evaluation. Also emphasized is the need for careful and explicit specification of the data-generating model before undertaking data analysis. (Author/JKS)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED143675

U S DEPARTMENT OF HEALTH.
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BLENDING EVIDENCE, TECHNIQUE, AND JUDGMENT IN EDUCATIONAL RESEARCH INFERENCE

James Fennessey

Department of Social Relations
The Johns Hopkins University
Baltimore, Maryland 21218

November, 1976

TM006 378

Blending Evidence, Technique, and Judgment in Educational Research Inference

BEST COPY AVAILABLE

FINAL REPORT

National Institute of Education Project No. 74-0071

James Fennessey

Department of Social Relations
The Johns Hopkins University
Baltimore, Maryland 21218

November 1976

The research reported herein was performed pursuant to a grant with the National Institute of Education, U. S. Department of Health, Education, and Welfare. Grantees undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

SUMMARY

For a variety of reasons, educational research reports often are less useful than they might have been. One contributor to this situation is the use of a largely inappropriate paradigm of research inference. The recently developed "Bayesian" paradigm of research inference appears likely to have the characteristics needed to resolve many of the present difficulties, but the prospects are not entirely clear.

This project was undertaken to explore the Bayesian paradigm as it can be applied to educational impact studies, particularly studies on the education of the disadvantaged. The project's purpose was to ascertain the feasibility of using this new paradigm, and to identify its particular advantages and limitations when applied to present-day educational research.

The work of the project has revealed that, although the basic ingredients to exploit the Bayesian approach exist, a large amount of development work is still needed before a mature computer programming technology will be available. Accordingly, one product of the work on this project is several computer programs. These programs, described in this Report, facilitate the application of Bayesian analysis to the sort of data most frequently encountered in educational evaluation. A second outcome of the project is the demonstration of some particular advantages of the Bayesian paradigm as a basis for representing knowledge when the data is "weak". This point is exemplified in the Report by a reanalysis of the summary data from one compensatory education study. Finally, a third result of the project has been

the focusing of awareness on the need for very careful and explicit specification of the complete data-generating model postulated. Such awareness can be facilitated by representing the models via a computer simulation program. A prototype version of such a program, focusing on certain currently problematic aspects of public education, is presented as an additional product of the research.

The main conclusion of the present project is that the Bayesian paradigm, if coupled with a clear and formal mechanism for specifying the exact nature of the process believed to be generating the outcome data, can serve as the catalyst to stimulate a major improvement in educational research. However, such an improvement will occur only if further technical development work is performed, and if the users of such a paradigm, as well as the consumers of educational research, are able to re-orient some of their basic views as to the nature of social science inference.

PREFACE

This report represents the fruits of an exploratory project, a reconnaisance outside the usual confines of conventional educational research. It was undertaken in a spirit of optimism, and is concluded with a somewhat different feeling, but a no less positive and encouraging one. The Bayesian paradigm of statistical inference was the focal point of this investigation, particularly as it might be of benefit to research in education. As a result of the work accomplished during the past two years, our optimism has matured considerably, and is now perhaps better described as sober confidence. The Bayesian approach is not a panacea; on the contrary, one major outcome of this investigation is that the paradigm has illuminated some otherwise obscure problems in the conventional practice of research design and inference. Yet, the problems today seem more manageable, because they have been recognized and delimited clearly. Moreover, as a result of the work described in this Report, we are now more prepared to make an intelligent attack on the real problems that plague educational studies, and in fact much social science research. This present study has been one of exploration; future work will be that of conquest and settlement.

Several persons deserve special acknowledgment for their contributions to the success of this project. Diana James, Marsha Paller, Susan Radius, and Jay Rockstroh each served at various times as highly capable assistants. Mr. Rockstroh's work in developing several computer routines was particularly valuable, all the more so in view of this investigator's very limited acquaintance with numerical methods. The staff of the Department of Social Relations, and that of the Homewood Computer Center at Johns Hopkins have contributed in many incidental ways to the facilitation of the work. Carlyle Maw of the National Institute of Education has been an ideal project officer, providing encouragement and useful leads to areas of knowledge seldom trodden by educationists. Finally, Ganie DeHart and Barbara Hucksall typed this final Report, performing this task with grace and efficiency under considerable time pressure.

In addition to these professional colleagues and coworkers, a personal acknowledgment is due -- more properly, a tangible debt is owed -- to my wife Gail, whose participation in the project was real and thorough, despite the fact that she is proudly a non-quantitative educator.

All of these persons have done their best to support this work, and I have done my best to carry it through. This document is indeed the Final Report of one project, and I hope it can truthfully be called a Progress Report as well. The challenge which provoked this project continues to exist, and we shall continue our efforts to deal with it. Those who have worked on the project join me in hoping that our readers will find this material of sufficient interest and value that they will begin to take part in the larger effort of overhauling and improving the way in which we approach educational research inference.

James Fennessey
November, 1976

TABLE OF CONTENTS

Page

Summary i

Preface iii

Chapter

1. Introduction 1

   1.1 The Problem 1
   1.2 The Bayesian Paradigm and its Promise 6
   1.3 Overview of the Report 10

2. Statistical Inference in Research 19

   2.1 Introduction 19
   2.2 Data-Generating Devices and Models 20
   2.3 The Gaussian Process 32
   2.4 Conventional and Bayesian Logics of Inference 40
   2.5 Scientific Research Methodology: Conventional and Bayesian 46

3. Implementing Bayesian Analysis Procedures 69

   3.1 Modeling the Academic Achievement Process 69
   3.2 Developing an Analysis Strategy and Procedure 72
   3.3 Developing a Prototype Analysis Package 78
   3.4 Programming the Student's _t_ and Behrens-Fisher Distributions 83
   3.5 Exploring the Utility of Numberical Integration 97
   3.6 Programming the Inverse Chi Distribution 111

4. A Bayesian Reanalysis of the Incentives in Education Data 125

   4.1 Introduction 125
   4.2 The O.E. Incentives Study 129
   4.3 Reanalysis from a Bayesian Perspective 136
   4.4 Reanalysis at the Level of Individual Students 143
   4.5 Reanalysis at the Level of Grades 171
   4.6 Analysis at the Level of Schools 180
   4.7 Sensitivity to Prior Beliefs 185

Chapter Page

5. Using Formal Models of Social Process 211

5.1 Introduction 211
5.2 The Classroom Dynamics Model - Background 214
5.3 The Classroom Dynamics Model - Specifics 216
5.4 Toward a Heuristic of Inference 223

6. Summary and Conclusions 261

6.1 Summary 261
6.2 Conclusions 263

References 266

1.0 Introduction

1.1 The Problem

On almost any specific topic in educational research, one can readily find inconsistencies in results between one study and another. This situation, (we believe) results (in large part) from the universal use of an inappropriate framework of inference in educational research. This prevailing paradigm for inductive inference within the context of statistical analysis also has contributed to a variety of practical problems. For example, educators who wish to use research evidence in making their practical decisions often become confused, discouraged, and even skeptical or suspicious. This situation constitutes a serious, widespread, and stubborn problem. It affects "the production and utilization of knowledge", as discussed in the National Institute of Education's 1974 program announcement, not only in education, but indeed in many kinds of social science research, including some work very remote from education. Although the basic problem is a very abstract and generally occurring one, it can be more fruitfully considered in a single fairly well defined context. The general aim of project reported here was to begin such a consideration.

The existence in the research literature of contradictory and confusing research reports on a single general topic -- such as those on the impacts of compensatory education programs -- constitutes a barrier to the utilization and advance of scientific knowledge. It directly reduces the credibility and relevance of research results as an input to the policy formation process. Perhaps more important for the long run than the specific controversies is that such a situation breeds a general mistrust of social research, and thus makes such research more difficult to perform successfully. Another consequence affects the researchers themselves. Like their counterparts in educational management, they frequently find the picture a confusing and chaotic one. Although this experience does not prompt them to abandon their commitment to research (at least not usually), it does substantially inhibit the progress of disciplinary knowledge about the education process. One consequence is that researchers become motivated to channel their efforts to areas where such confusions will be reduced. But this in turn means that efforts are further diluted, and the general consequence is the creation of an ever-expanding web of research -- perhaps analogous in some ways to a communal spider web. The research film is spread over a very wide area, and occasionally does catch a fly (i.e., finding) which is then quickly devoured by the spiders. Most of the time, however, the spiders spend their time simply weaving more strands, gradually covering a broader area. But, the web is never developed and strengthened to the point where it can capture and retain any object of substantial size or vigor. To the educational practitioners, it appears as an occasionally interesting distraction, but seldom a real source of aid.

Building on a gradual personal recognition of this communication-cumulation problem, as it can be called, this author began a search to locate and evaluate alternative frameworks for inference that might be used in education research. Any such better framework would have to be one which can reveal rather than mask the sources of contradiction between studies. Also, it would have to report the results of research in a more intuitively meaningful form than those now in customary use. Finally, a better framework should be one which positively facilitates the cumulation of scientific knowledge about education.

The research project reported here, therefore, was not derived by the usual process of examining prior studies in a single substantive area where knowledge is presently lacking. Likewise, it did not evolve simply from practical administrative experiences in a single kind of education setting -- although there was abundant experiential corroboration of the need for it.

Instead, the present project was derived from an informal analysis of the current patterns of communication and influence in the field of educational research. Some of these patterns involve contacts between practicing educators and educational researchers; others involve communications among different researchers. The primary form of communication considered has been the research report and commentary on reports. Our review of such reports and their utilization indicated that many of them either mislead, mystify, or confuse their readers. They mislead other researchers by conveying an impression of solidity and strength which is sometimes greater than warranted, other times less than warranted. They mystify the professional educator because they are couched in an unfamiliar technical jargon, and because they employ a logic of inference which is cumbersome and circuitous. Finally, they confuse both kinds of reader because the generalized conclusions from one research project on a topic frequently are contradicted by other evidence -- either by other reports on the same topic, or by the reader's own experiences. These contradictions all too often are not resolved by research specialists, and simply remain in plain sight as enduring puzzles.

In carrying our preliminary analysis further, it became clear that this "internal contradiction" aspect of research reports arises mainly from two sources. First, in most educational research situations, the researcher and the readers of his report each have rather specific and firm initial beliefs about some aspects of the phenomenon under study. These need not be beliefs about what the specific outcomes of the research will be, but usually they are beliefs which do affect the design and implementation of the project, and the interpretation of the results obtained. Second, in a large proportion of educational research studies, the data actually obtained are such that they might have been generated with equal ease by any one of several distinct and different processes or mechanisms. That is, the data actually obtained from the studies do not greatly change the degree of plausibility a reasonable person would attach to each of several alternative explanatory models for the situation under study. In other words, the data are low in diagnostic value.

This analysis of the communication problems prevalent among educational researchers and their practitioner audiences has suggested several things. First, any effort to improve matters should be one which draws attention to the content and the firmness of the assumptions being made by the researcher; it should show as fully as possible the way in which these assumptions affect his interpretation of the results. Second, the method also should provide a clear indication of the diagnostic power of the data with respect to the questions being addressed, and the alternative explanations under consideration. Finally, any new method of analysis and inference should be able to improve communication by sharpening the articulation possible between separate studies on related topics, and by facilitating the cumulation of knowledge in an area as more and more investigations are done. The presently informal and laissez-faire mechanisms for such cumulation have been notably lacking in success.

If a new, more formalized and more direct, framework for research analysis and inference can be created and used for reporting and discussion, and if this new framework can transform the patterns of communication among researchers and practitioners, the benefits to both groups will be enormous. Practitioners will benefit through being freed from the distraction and confusions of useless yet intriguing research results. They will begin to understand, believe, and actually to use the research evidence which is relevant to their particular specialized sphere of action. At the same time, researchers will be freed from participating in excessive, empty, destructive controversies; they will no longer have to confront the many paradoxes of persistent inconsistent findings as mysteries. Moreover, researchers will more readily be supported in their efforts to develop and to execute research projects which are simply better. That is, their designs will be more conclusive, their measurement strategies more powerful, and the articulation between their results and the plausible patterns of explanation will be more explicit and tight. In other words, if an awareness develops as to the low diagnostic value of much currently-produced research evidence, there will emerge as a consequence an acceptance of -- even a strong demand for -- research designs and procedures which can be expected to yield results of real practical and disciplinary value.

1.2 The Bayesian Paradigm and its Promise

During the past twenty years or so, as this communication-cumulation problem in education research, and indeed within much of social science research, has been growing, important developments have been occurring in other intellectual fields. Some of these developments, particularly in the areas of philosophy of social science, mathematical statistics, and management science, appear to offer the ingredients for an attractive solution to the problems just described. Within the discipline of mathematical statistics and of data analysis methodology, the so-called Bayesian paradigm for statistical inference and for scientific induction has taken shape and has evolved rapidly. Our initial inspection of this paradigm indicated that it appears to possess all of the general characteristics necessary to deal with the problems outlined above. Also, it has several other desirable features, but some specific limitations and areas of uncertainty as well. Thus, the present project was devised to undertake a serious consideration of the Bayesian paradigm of inference as an alternative framework for the furtherance of educational research.

Interestingly enough, the general ideas of this Bayesian approach are not recent. In fact, the basic work from which the paradigm finds its name was done by Thomas Bayes in the 1750's and published in 1763. The Bayesian tradition continued along as a major theme in statistical inference and analysis until the later 1800's and the first half of the twentieth century. At that time, two intellectual developments combined to make the Bayesian approach less popular. A general emphasis on operationism and objectivity among scientists made the Bayesian view of probability unappealing to them and led most scientific methodologists to prefer the so-called frequentist approach propounded by Venn, Von Mises, and (more recently) Reichenbach. Second, the enormously creative work of data analysis geniuses such as Karl Pearson and R. A. Fisher, and the maturation of the Neyman-Pearson paradigm -- which conceptualizes statistical inference in terms of sampling distributions -- caused this "sampling-theory" paradigm to flourish within scientific data analysis. Another advantage of this paradigm was that it was easily applied to real world scientific problems using only a moderate degree of computation power. The Bayesian paradigm, conversely, was perceived by scientists as excessively subjective and abstract in style, and excessively difficult in computations for any but the simplest real-world problems. Thus, as the developing thrust of statistics became motivated not by study of games of chance, nor by deep philosophical questions about belief and action (cf. Hacking, 1975) but instead by the needs of practical scientists, typically in agricultural and biological work settings, the Bayesian viewpoint on inference sank into obscurity.

But, starting in the 1920's and 1930's resulting from the work of F. P. Ramsey, J. M. Keynes, B. DeFinetti, and then continuing in the 1940's and early 1950's with further work by H. Jeffreys, R. Carnap, I. J. Good and L. J. Savage, the Bayesian inference tradition (in two or three main variants) has been revived. In the later 1950's and since, however, the general ideas and the technical machinery of this Bayesian perspective or paradigm of inference have been developed and studied almost entirely apart from educational research. Basic work was initiated and advanced by mathematical statisticians such as DeFinetti, Jeffreys, Savage, and D. Lindley. The Bayesian approach was fleshed out and applied in practice mainly (though not exclusively) by specialists in the quantitative analysis of business and other management decisions. The leading workers in this respect have been Howard Raiffa and Robert Schlaifer of the Harvard Business School. In their work, they have married the Bayesian inference perspective very closely with the formal logic of decision analysis, itself a recent intellectual creation springing largely from seminal work by VonNeumann and Morgenstern on the branch of mathematics known as game theory.

At present, the vast majority of educational researchers are hardly aware of the existence of the Bayesian paradigm for practical statistical inference. Fewer still have seriously attempted to use it. One noteworthy exception to this has been the work of Melvin Novick and his associates at Educational Testing Service and more recently at The University of Iowa. Novick has written papers on the foundations of Bayesian statistical inference in statistics journals, has co-authored a statistics text stressing this approach, and has made some applications of the Bayesian ideas in predicting academic performance as part of an admissions decision. Like Robert Schlaifer at Harvard, he has also been the guiding spirit behind the development of one package of computer programs for carrying out a Bayesian analysis.

However, despite Novick's work, and despite fairly widespread use of Bayesian statistics in certain areas of management science and engineering statistics, there has been little application of the Bayesian ideas in educational research to date. In fact, although a few articles discussing the Bayesian approach have appeared in the literature of psychology and sociology, it seems accurate to say this is still a very rare approach within the social and behavioral sciences. Perhaps one reason for this is the insecurity of social researchers about really being scientists after all. Imitating the chemists and biologists, social science workers seem to have neglected the work of professional practice experts (business, medicine, engineering) and to have continued their thinking within the sampling-theory framework of inference. That framework has indeed been fruitful and powerful, and yet recently the results, broadly considered, from it have been disappointing -- as discussed in the earlier section. We will argue later on that the "reach" of our collective curiosity about social phenomena generally, and in particular about educational phenomena, over the last ten to fifteen years, has increased dramatically. This increase has been so great, so qualitative, that the "grasp" provided by conventional statistics is now exceeded by that reach of curiosity (cf. Fennessey, 1972). But, on the basis of its analytic character, and particularly if embedded in a broader general philosophy of research inference, the Bayesian paradigm appears as a promising alternative framework.

1.3 Overview of the Report

In Chapter 2, a fairly detailed presentation and discussion is given of the logic of statistical inference in a research context. A number of key terms are defined and illustrated, and some of the foundamental features of the Bayesian paradigm which distinguish it from the more conventional paradigm are discussed. That chapter deals not with education phenomena *per se*, but with the logic of inference. Although the presentation there is direct and concrete as much as possible, the ideas are subtle. That is, the chapter may appear at first to deal mainly with the commonplace, but in fact is aimed at making some fairly unobvious but important points about the characteristics -- logical and practical -- of the two frameworks of inference.

The succeeding chapters have been organized to present, as nearly as possible, the results of our work on this project in correspondence to the list of objectives given in the original proposal. A brief discussion of those objectives, and our success in meeting them, is presented in the remainder of this section.

The first objective stated in the project proposal is that of providing guidance as to which of the well-developed Bayesian model formulations (e.g., the binomial, multinomial, or normal/inverse chi) appears most appropriate for application to studies of education program impacts. Since this type of study not only is interesting in its own terms, but also is a prototype for "treatment impact" studies in a wide variety of substantive contexts, the application of an appropriate Bayesian formulation to it can be especially useful. Our work has led to a dual conclusion. First, the "normal/inverse chi" formulation is the most typically appropriate model. Second, and somewhat frustrating, work with this model (and particularly with the Behrens-Fisher specification of it) has been quite limited. In fact, much more work has been accomplished in developing and illustrating the binomial model as an example of Bayesian inference. Incidentally, one of the general findings of our work on this project is that the Bayesian inference viewpoint has been much more often discussed and studied in terms of comparing specific discrete hypotheses (each of which has a numerical probability of being applicable, before and after the data has been analyzed), than in terms of estimating prior and posterior density curves for the continuously variable parameters of some generating process. But, in any case, our principal result with respect to this objective is that the specific formulation most suited for attention and development is the "normal/inverse chi".

The second objective stated in the proposal is that of providing some practical information on the consequences of choosing different prior probability assessments. This can be described more specifically as an analysis of the sensitivity of conclusions to choice of prior belief densities about the numerical parameters determining the outcomes of the process. Our results on this question indicate that, in comparison to the traditional methods, the Bayesian approach for impact analysis provides enormous flexibility. Thus, it has been noted that in the normal/inverse chi context, the prior probability assessment for the process mean can be varied in three distinct respects. That is, one's beliefs about the process mean can be (and must be) formalized with respect to the location, the spread, and the confidence they reflect. A discussion of this part of the analysis is given in the chapter which reports the results of applying the Bayesian approach, in the form of the normal/inverse chi model, to data from the U.S.

Office of Education's 1972 study titled "Incentives in Education".

The third objective mentioned in the original proposal is to investigate in what ways and in what degrees the results of any research analysis are dependent on the substantive aspects of the data generating model. This has proven to be a fairly complex and difficult issue. We have a fair amount of work and progress to report, but more investigation is needed. In pursuit of this objective, it became clear that the model specification activity needs to be subdivided into at least two parts. First, there must be specification of what can be called a proximate model, and, second, there must be specification of what can be called a global (or contextual) model. The proximate model typically contains some specific parameters in a statistically tractable form (and in our case has thus far always been the "normal/inverse chi" model). This proximate model is, however, necessarily seen as embedded in a more complex and comprehensive "global" model of the social phenomenon under study. It has become clear that formalization of this global model is at least as important for a confident Bayesian inference within the context of educational research as is the assessment of prior beliefs about the parameters of the proximate model. Accordingly, one objective of the project which emerged in the course of doing the work was the development of a prototype for such a global model formalization. Specifically, we have developed the core of a computer package for simulating the educational achievement process. This package is described in detail in the chapter on modeling classroom dynamics.

Incidentally, it is worth noting that our sensitivity analysis has emerged as requiring, so far, three distinct facets. These are (1) sensitivity to the form of the proximate data-generating model; (2) sensitivity to the choice of prior beliefs about the parameters of that model; and (3) sensitivity to the content and relationships contained in the global process model. Thus, we have coined the term "multi-faceted sensitivity analysis" to describe this finding.

The fourth objective described in the original proposal concerns carrying out an investigation of the quality of some already collected data, in terms of its ability to distinguish between alternative explanatory models. This data was to be found in several typical studies of educational program impact. As events have turned out, this objective has been achieved only in an indirect and limited way. Only one study has been explicitly re-analyzed, due to the need to devote greater than anticipated amounts of effort to the more general methodological topics. In that re-analysis, however, one result emerging with clear focus is that the question of data diagnosticity is closely tied to the generally neglected question of the proper level of aggregation. That is, whether the data is valuable or not depends in large part on whether a study is seen as dealing with individual students, or with classrooms, or with grades, or with schools. Unfortunately, most educational impact studies in fact are concerned with units more aggregated than individual students, and this drastically limits the potential diagnosticity of the data. When the relevant units are grades or schools, then the number of cases studied is often quite small in comparison with the number of cases any practitioner might choose as a reasonable estimate of the worth of his prior beliefs (this notion ties us to the question of conjugate forms of prior distribution, which will be discussed in a later chapter).

Related to this fourth objective is the study of reasons for apparently contradictory results. This topic likewise has proven more complex than originally expected. However, the work done to date has created a substantial set of resources (several analytic programs, and the classroom dynamics simulator) which can be employed in future investigations.

The fifth and last objective stated in the proposal for this project was that of investigating and evaluating the specific computational and practical resources now available and difficulties encountered. A good deal of our research effort, considerably more than originally planned, was unavoidably devoted to this topic. Our work has revealed that the presently available computational aids, though in principle nearly adequate, suffer from a number of major and minor limitations. The minor limitations concern simply program organization and operating convenience. Although these weaknesses do cause sizeable problems for a researcher seeking to analyze actual data, they could be rather easily dealt with by the application of careful design and programming procedures -- such as those characterizing the present generation (but not earlier ones) of programming packages for orthodox statistics (e.g., SPSS, P-Stat). The more fundamental issues revealed by our work concern the appropriate approach, within numerical analysis considerations, to handling situations where numerical integration, and even multiple numerical integration, must be done. We have some tentative recommendations on this point. Our findings and recommendations also concern the question of how best to deal analytically with the fact that distributions constitute the basic input and output of a Bayesian analysis.

On the numerical integration topic, we have investigated several general approaches within the context of evaluating the Behrens-Fisher distribution, and have some findings about the comparative advantages of Gaussian quadrature, Romberg integration, and a modified Simpson's rule algorithm. We find the modified Simpson's rule formula generally preferable for this situation. We also offer some general ideas worthy of further investigation with respect to multiple integration. Little is known (at least in the literature typically available to social scientists) about these topics, and in particular there is little communal sense of the actual tradeoffs between computing cost and needed accuracy in the context of Bayesian analysis.

With respect to the question of handling entire distributions analytically, we have recognized once again the applicability of the phrase "there is no such thing as a free lunch". Just as a distribution has advantages in being easy to present and understand pictorially, so it has disadvantages if one wishes to make further analytic use of its content. This point remains true whether the distribution in question represents a prior, a likelihood, or a posterior. It is likewise true that difficulty arises whether one wishes to use the distribution analytically as an input to a model for sensitivity analysis, or as a dependent variable in such an analysis. Fundamentally, a distribution is a complex object, not easily reducible to a single number. This complexity interferes considerably with any systematic sensitivity analysis. Two general approaches appear worthy of further study. First, one might use the basic idea underlying Gaussian quadrature as well as much classical numerical analysis, and consider substituting for any distribution several summary parameters derived from (orthogonal) polynomial approximations of the original distribution curve. Second, one might instead simply talk of analyzing the distribution in terms of its moments (e.g., a mean and variance). Here again we are implicitly using a polynomial as an approximator. This topic too deserves further investigation by a team which includes mathematical as well as social science expertise.

So then, our conclusions and recommendations will cover substantial ground. They deal partly with data analysis methodology in a fairly narrow and technical sense, but also with the logic of educational research in a more general sense. We have explicated rather fully the idea of a multi-faceted sensitivity analysis, and regard this as an important methodological concept. We have pointed out that such analysis can and should address itself to at least the following facets: the global model; the proximate model; the various aspects of the prior belief assessments involved in the proximate model; the data quality; and choice of units of analysis. Our work on this project has led to the development of these findings and conclusions and to the creation of several specific items of technology necessary to pursue this new and more powerful approach properly.

We remain highly optimistic about the Bayesian paradigm on the basis of our findings, and regard as the main contributions of the project (1) the clarifications it has provided as to the features and limitations of the paradigm in an actual social science context; and (2) the specific steps that have been taken toward meeting certain research technology needs. Also, the project has promoted the development of a clear agenda for work toward the accomplishment of further steps. It is our judgment that with a reasonable degree of development effort, this new paradigm could become very widely and successfully utilized within approximately five to ten years. Before this will be possible, however, much foundation-laying work will be needed.

2.0 Statistical Inference in Research

2.1 Introduction

In talking with a number of educational researchers and other social scientists, it has become clear to me that many of the ideas which are basic to the Bayesian viewpoint on inference--and indeed even ideas (such as that of a formal model for a data-generating process) which are basic to all varieties of scientific statistical inference--can be quite unclear and mysterious. As a result, despite the fact that a person is well-acquainted with statistical terminology and the mechanics of data analysis, these ideas can fail to be a part of his or her appreciation of the research inference logic. A sizeable number of research producers, and a still larger fraction of research consumers, learn to go through the proper motions without ever developing a proper appreciation for the deeper meaning of what they are doing. Formulation and analysis of a problem in terms of a statistical model is, in short, often done more by rote than by rationality.

Since this is the case, any improvement in payoffs to be gained from research which uses the Bayesian framework can be expected only if these essential ideas are communicated more vividly and forcefully. Ways must be found to improve the demonstrations and illustrations of the really basic ideas in statistics. This need for an improved pedagogy is especially severe among those learning social science research methods. A proper balance must be struck between development of fundamental ideas of scientific inference and the discussion of statistical formulas.

As a first attempt to provide this discussion, the present chapter describes some general ideas necessary for an understanding of the Bayesian framework for inference, and for an appreciation of the ways in which it differs from the classical framework. An attempt is made to provide a balance between ideas underlying the logic of data analysis, and the discussion of probability formulas and statistical inference.

2.2 Data-Generating Devices and Models

The single most central idea in applied statistics, and in scientific research inference, is that of a formal data-generating model for a particular phenomenon. Such a model exists in one sense inside the investigator's head, but in another sense it exists (because it is formal and part of the scientific enterprise) also as part of the specialized subculture of scientists interested in that phenomenon. Insofar as it is formal, such a model consists of a set of mathematical expressions, involving constants, parameters, and variables. Depending on the kind of phenomena and the situation being studied, the data-generating model associates a probability value or a probability density with each possible value of a particular focal variable in the research. The focal variable is regarded as being stochastic in nature. That is, at least as far as our knowledge of it is concerned, exact values of the variable cannot be deduced from any set of premises; however, there is regularity in the pattern of its occurrences. This paradoxical combination of individual irregularity and aggregate regularity is the distinctive feature of stochastic (sometimes called random) variables.

The discipline of probability theory investigates the patterns of observed data to be expected, when the general form of the model and the numerical values of the parameters are given. The discipline of statistics, and especially that of statistical inference, investigates the "inductive" problem of knowing what parameter values are most likely, given the general form of the model and some actual set of observed data. This inductive problem is at the heart of scientific inference as well, and has been a difficult one for centuries, both for philosophers and also for each new branch of science. The difficulties have typically resurfaced in the wake of each new expansion of scientific ambition and scope. In this report, we can only adumbrate some of the complexities of this problem of induction in science, but we shall try to indicate that it is at the core of the communication-cumulation problem which prompted the present program of research. Incidentally, notice that both forms of inference--the deductive inference typical of probability theory and the inductive inference typical of statistics--take as a premise the explicit form of the data-generating model to be applied in the particular situation. In actual scientific practice, of course, choosing an appropriate model for a new area of investigation is the heart of the challenge confronting science. An extremely illuminating discussion of this choice aspect of scientific work is provided by Ian Hacking (1975) in his description of three distinct data-generating models currently used in physics, each applying to some phenomena but not others (Maxwell-Boltzmann statistics, Bose-Einstein statistics, and Fermi-Dirac statistics). This illustration makes clear that the scientist (or the administrator) must use his own best but tentative judgment in choosing a data-generating model whose general form appears suitable for the particular class of situations under study. One of the interesting features of such modeling in science is that after a suitable model has been created and empirically validated, there is usually some difficulty in imagining how the other models could ever have been seriously considered; their weaknesses are quite obvious. The fact neglected here, of course, is that the currently-known properties of the phenomenon that support one model and contradict others were strictly unknown until scientific investigation turned to the comparison of the several models.

Statisticians speak of the "model" or the "process" routinely, usually without being very explicit as to the meaning, the centrality, and the problematic character of those terms for actual scientific work. This is a perfectly legitimate part of the division of scholarly labor, but it places an additional burden on those who seek to upgrade the level of research practice.

To clarify what is meant here, and to concretize the discussion, it is desirable to begin by speaking about a concrete data-generating device, or a data-generating mechanism, rather than the abstract and strictly symbolic idea of a data-generating model. Such a device or mechanism is sometimes referred to as a "black box" among scientists. Nothing is assumed known about the workings inside the box, but it does manifest activity by generating things that are taken as data. More generally, the idea is that one has a device which accepts certain inputs and produces certain outputs, but that no knowledge is available as to the internal relationships that produce the outcome. For instance, to be really concrete, we might visualize a small device something like a pocket calculator. This machine will have a display register which can show a several-digit number, and a single button on the front. Each time the button is pressed, there is a slight pause and then the display shows a particular number. Each time the button is pressed, the number produced is almost always distinct, and there is no obvious pattern in the numbers over the sequence of pressings.

Such a machine clearly is a data-generating device, in fact one which directly generates quantitative data, but it is somewhat mysterious. It is a "black box". That is, there is no apparent way an observer can gain greater understanding of exactly how the number in the display was generated. In other words, we have a concrete data-generating device, but do not have any corresponding explicit abstract model for it. In the absence of further knowledge of the device's workings, the only kind of model for it we might construct will be one which chooses digits at random, and combines them into numbers. Thus, we would regard the device being modeled as stochastic, without intending to take a position as to the real nature of the mechanism. If the stochastic model reproduces the behavior of the device in all ways in which we are interested, or if it does fairly well and no better model can be proposed, then we will continue to use that stochastic model for this device, regardless of whether the true device is stochastic or not. This same logic--that is, choosing a purely random and "equal-chance" model when no preferable alternative is available--has had a long and controversial history. It has been variously known as "Bayes's postulate", LaPlace's principle of insufficient reason, the principle of indifference, and the strategy of the uniform prior. We shall not develop the elements of this controversy now, but do wish to note that in many investigatory contexts it does seem sensible and not particularly risky to employ such a model, particularly if the employment is consciously tentative and subject to revision in view of further evidence that may become available.

Now, let us consider instead of our calculator-like device, one which is simpler, and one whose workings are immediately apparent upon inspection. A simple "spinner" is such a device. That is, we take a circular piece of stiff cardboard, make a small hole in the center, and then insert a pin through the hole. Then we attach to the pin a small plastic arrow, so that the arrow can spin freely around and point to any region of the circle. Finally, we mark off one fraction of the circle's circumference by drawing a shaded border along it, and leave the remainder of the circumference white. This device now corresponds to the classical notion of a chance device with probabilities determined by shape and symmetry. In other words, we feel confident on the basis of much similar experience that if the spinner is struck and spun, it may stop pointing in any direction at all, and that all directions are equally likely (assuming for the sake of the argument that the construction of the device is uniform).

If the arrow is spun, it makes several revolutions, and then slows and stops. If it stops while pointing to a shaded portion of the rim, then we define the outcome as being +1; if it stops over the white portion of the rim, we define the outcome as being 0. Thus, we have attached to the outcome a random variable (a datum) which can take on only two values, +1 and 0. Also, we have a concrete and understandable device by which values of that variable can be generated. Such a spinner produces outcomes in a way we regard as probabilistic, or stochastic, or random. We feel that this characterization is somehow more objective than the previous characterization of the number display machine. But the difference is not intrinsic to the two machines or devices. Rather, it concerns our familiarity with devices of this kind, and the kinds of uses we would hope to make of the resulting data. Devices such as the spinner are usually thought of as "chance devices" simply because the manner of their construction and use clearly makes it very difficult to predict the outcome, so difficult that it has never been seriously attempted. But, we would tend to suspect that the number display device described earlier has some kind of deterministic formula by which it generates numbers, and we might try various possible explanations to see if they seem to work. In other words, we recognize that in constructed devices there may be some which should be studied and unraveled, and others which simply should be regarded as chancy or stochastic.

The point of this somewhat elaborate discussion is simply that a similar openness toward thinking in stochastic terms is needed in science. If a phenomenon is being addressed, it is entirely reasonable to ascribe all or part of its variation to chance. It is a matter of complex judgment, taking into account a number of circumstances, whether any means to understand further the mysterious aspects of it is available (cf. the Heisenberg uncertainty principle) or in any event whether the implied effort is worthwhile. Just as the spinner might in principle be understood, measured, and tracked so that its outcomes were predictable--or at least more predictable, so the same thing is true of many scientific phenomena. But, it is also true that the construction of the spinner (or its more sophisticated version, the roulette wheel) and the circumstances of its customary use both make such further knowledge infeasible and not of general utility. This too has parallels in scientific contexts. Wesley C. Salmon has a good discussion of some related ideas about chance and science in his 1971 book.

Let us now consider the spinner somewhat more closely, and attempt to "understand" this concrete device more formally by constructing a corresponding abstract model for the generation of this same kind of data. Upon introspection, we realize that we believe that the spinner is so constructed that the outcome on any one trial is independent of (unrelated to) the outcome on any other trial. That is, the outcome on trial k is not known any better by knowing the outcome of trial k-1, or in fact by knowing any series of outcomes from other trials. Again, we realize that we have been assuming that the same amount of the rim of the circle has been shaded on each trial. Now, by using some of the axioms of probability, we develop the formal equation for the probability of various sets of outcomes over a series of trials, particularly the number of times the spinner points to the shaded portion. We take as nonproblematic features of the model the number of trials to be considered and the amount of the rim that is shaded. We express the formal model by an equation using symbols such as N for the number of trials, p for the proportion of the rim that is shaded, and S for the number of times the arrow points to the shaded portion. But notice that we can also dispense with the idea of an arrow and a circle. We have simply a model where there are two possible outcomes, labeled 1 and 0, and the probability of outcome 1 on any trial is p and is independent of any other trials. This model is called the Bernoulli process model for binary data. Notice that in fact the model we are describing is a whole family of models. There is a distinct model for each possible value of p, and for each possible value of N. What those models give is the probability of each possible value of S (where S can range from 0 through N), conditional upon specified p and N. Notice too that even with this model, it is not possible to make exact predictions about the outcome of any single trial or any series of trials. In fact, we have rephrased our interest so that we are concerned only with the total number of 1's in the series of trials. And, even for this sum, we can state our knowledge only in probabilistic terms--the probability of various values for S.

Thus, when we have specified the form of the model,

$$\text{Prob}(S) = \binom{N}{S} p^S (1-p)^{N-S} \qquad \text{for } 0 \leq p \leq 1 \quad \text{and} \quad 0 \leq S \leq N$$

and in addition have specified numerical values for p and N, then we can construct a probability relation between S and the probability of S. Following a convention used in econometrics, we shall refer to a particular model with all necessary parameters specified numerically as a data-generating structure. Thus, the abstract equation is a model; the equation with all parameters numerically known is a structure.

So, given p and N, we would summarize our knowledge about S by reporting (perhaps via a plotted curve) the function relating all possible values of S to their probability of occurrence. We would state this in mathematical terminology by saying that the probability of S is a function of S, with the parameters p and N appearing in the functional formula. We shall return to this notion of the nature of probabilistic knowledge in a little while. Incidentally, although we shall use the term "structure" sometimes to indicate the formal, specific equation that constitutes our belief about the data-generation device, we shall also use the word "process" in a slightly more generic sense to refer either to the actual real-world phenomenon, to the abstract model, or to the specific structure. This usage will not create any problems in context.

Alternatively, we might consider a different kind of data-generating device or mechanism, though one resembling in some ways the one we've just been discussing. The rim of the circle might be divided into a large number of equal-length segments--perhaps 100 or even 1000--instead of just two segments. Thus, if we assume that there are 100 segments on the rim, and the arrow is spun, the possible outcomes might be defined as a variable which takes values of 00 through 99 inclusive. It should be apparent that each of those 100 outcomes is equally likely, due to the design and construction of the device. If this new device were cycled through a number of trials, those outcomes could be modeled by a multinomial distribution, which is a generalization of the binomial used in the Bernoulli process. Now, if we extend the idea so that there are an infinity of points along the rim of the circle, and if we assume the circumference of the circle is 3.1416, then we would have an abstract model in which the outcome follows what is known as the "uniform" distribution on the interval (0, 3.1416). This new outcome is a continuous random or stochastic variable. It may,in other words, take on any value between 0 and 3.1416. There are infinitely many such values. Thus, we must not speak of the probability of a particular outcome, but instead of the probability density at any particular point in the interval. Despite this conceptual change to match the shift from discrete outcomes to continuously variable outcomes, the general logic of the discussion is almost entirely the same. Our knowledge about the outcomes is still expressed by a function relating each possible outcome to the probability (density) of its occurrence. That density function for the uniform distribution, however, has only one parameter, the length of the interval within which the outcomes can occur. In this example, the parameter is 3.1416. Thus, in this case with a continuous outcome, we describe the probability form in a very simple way:

$$\text{Prob}^*(x) = \frac{1}{3.1416} = .3183 \quad \text{for} \quad 0 \leq x \leq 3.1416$$

where Prob* denotes
a probability density

These two process outcomes, the S from the Bernoulli process and the x from the uniform process, are called stochastic variables. Our knowledge about them is expressed in terms of a curve in which the X-axis shows all possible values of the variable, and the Y-axis shows the probability (or probability density) of its occurrence. We can and do use this framework to represent our knowledge even if in fact the actual real-world process may be deterministic. Thus, in the example of the number-display box, for a given state of knowledge, we are speaking relative to certain available evidence and representing the kind of model we feel comfortable in using. If the total available evidence supports nothing more specific, then the only statements we can make regarding the outcomes of a process may be those involving distributions. This distinction between probability distributions used epistemically and those used fundamentally is sometimes a confusing one. In each case, our model of the real-world mechanism is a stochastic one, but our reasons for adopting the model are different. As pure scientists, we should always maintain some ambition to reduce further the amount of chance present in our models, but this ambition must be pursued intelligently and efficiently, with a proper recognition of the value of chance as a conceptualization aid.

The more important notion here, however, is this viewpoint about the nature of our knowledge of complex facts. Our tendency to use the idea of a curve relating possible quantitative outcomes to their respective likelihoods of occurring is the underpinning of the Bayesian approach to inference. Just as in the discipline of probability we find ourselves using a curve to express ourselves when we are given parameters but the outcome values are stochastic, so in the opposite case, namely when we are given the data values and seek to describe our knowledge about the possible values of the parameters, we should also use a similar curve to describe our knowledge.

In summary, then, we have gone into some detail to explain the meaning of "knowing" an outcome variable when it is stochastic. We have stressed that such knowledge, though imperfect, is nevertheless consistent and useful. More important, perhaps, such a form of knowledge is unavoidable. We should therefore be prepared to conceptualize knowledge of something which is quantitative and may take on various values, on which we have some empirical evidence, and about which some formal relations are assumed, as also expressible in the same formal way--i.e., probabilistically or stochastically. But this is exactly the question of inductive knowledge of a parameter, and is the core of statistical inference.

Thus, the argument is that our "estimate" of such values should take the form of a curve, just as does our probabilistic knowledge about the outcomes. In fact, there is no cogent reason why the two kinds of knowledge should be treated differently. Objections to this approach arise simply from a reluctance to accept the idea of stochastic variation and epistemic probability at all, but cannot rationally be used to distinguish between one form of knowledge and the other. To paraphrase, we have described probabilistic deduction and suggested that if it is accepted, then so likewise should probabilistic induction. If our knowledge of variables can be expressed via probabilities, then so too can our knowledge of parameters. Interestingly, even the conventional sampling theory statisticians accept such an idea to some degree when they speak of the likelihood function. They do not, however, follow that concept's implications through consistently, nor exploit the advantages it makes available through application of Bayes's Theorem.

2.3 The Gaussian Process

But this gets us ahead of our argument somewhat. Let us return to the more specific discussion of forms of model. A less intuitive model, but a very important one, is the Gaussian or Normal distribution. A mechanism for generating data which follows this distribution can fairly easily be constructed. In fact, the kind of mechanism which will be described in the next few paragraphs is important in three ways: (1) it mirrors a large number of interesting real-world phenomena; (2) it provides a concrete analogue of the central limit theorem; and (3) it forms the basis of one widely used algorithm for actually generating via a computer program artificial data which follows the normal distribution. The attached computer program and output, Figure 2.1A and 2.1B, (written in BASIC for the Wang 2200) illustrate the approach.

The key to this straightforward mechanism for generating Normally distributed data is that the number which comes out of the mechanism as the "observed" value is in fact equal to (or a linear function of)

Figure 2.1A

```
1010 REM GENERATE NORMAL DISTRIB VIA SUM OF BINOMIAL VALUES
1020 DIM X(200,2),F(31)
1030 U=.184373:R4=4
1040 REM N = # OF NORMAL DISTRIB VALUES
1050 REM K = # OF TERMS IN BINOMIAL SUM
1060 REM 01 NOV 76  1212 VERSION
1070 PRINT :PRINT "NORMAL DATA GENERATOR"
1080 INPUT "N=, K=, P=",N,K,P
1090 PRINT :PRINT ,"N = ";N
1100 PRINT ,"K = ";K:PRINT ,"P = ";P:PRINT
1110 PRINT :PRINT "  COUNTER    NORMAL DEVIATE    BINOMIAL SUM"
1120 FOR I=1 TO N
1130 Z=0
1140 FOR J=1 TO K:Z=Z+INT(FNU(U)+P):U=FNU(U):NEXT J
1150 Z1=(Z-K/2)/SQR(K*P*(1-P))
1160 S=S+Z1:S2=S2+Z1↑2
1170 PRINTUSING 1330,I,Z1,Z
1180 X(I,1)=Z1:X(I,2)=Z
1190 NEXT I
1200 FOR I=1 TO 10:PRINT :NEXT I
1210 PRINT "   MEAN              VARIANCE         SIGMA"
1220 M=S/N:V=(S2-N*M↑2)/(N-1)
1230 PRINTUSING 1330,FNR(M),FNR(V),FNR(SQR(V))
1240 PRINT :PRINT :PRINT "****":PRINT "CUMULATIVE DISTRIBUTION"
1250 MAT F=ZER:T=0
1260 FOR I=1 TO N:F(X(I,2)+1)=F(X(I,2)+1)+1:NEXT I
1270 PRINT "DEVIATE","FREQUENCY","CUM FREQ","CUM PERCENT"
1280 FOR J=1 TO K+1:T=T+F(J):Z1=((J-1)-K/2)/SQR(K*P*(1-P))
1290 PRINT FNR(Z1),F(J),T,FNR(T/N)
1300 NEXT J
1310 PRINT :PRINT :PRINT "END OF PROGRAM":PRINT
1320 STOP
1330 % -###.####        -##.####        -##.####
1340 DEFFNR(X)=SGN(X)*INT(ABS(X)*10↑R4+.5)/10↑R4
1350 DEFFNU(U)=U*2592221-INT(U*2592221)
1360 REM END OF PROGRAM
```

Figure 2.1B

```
NORMAL DATA GENERATOR

                    N =  200
                    K =  30
                    P =  .5


  COUNTER      NORMAL DEVIATE     BINOMIAL SUM
    1.0000          -0.3651         14.0000
    2.0000           0.3651         16.0000
    3.0000          -0.7302         13.0000
    4.0000           0.0000         15.0000
    5.0000           0.7302         17.0000
    6.0000          -0.3651         14.0000
    7.0000           0.0000         15.0000
    8.0000           0.7302         17.0000
    9.0000           0.3651         16.0000
   10.0000          -0.7302         13.0000
   11.0000           1.0954         18.0000
   12.0000          -2.1908          9.0000
   13.0000           0.3651         16.0000
   14.0000           0.3651         16.0000
   15.0000          -0.7302         13.0000
   16.0000          -0.3651         14.0000
   17.0000           0.7302         17.0000
   18.0000           1.4605         19.0000
   19.0000          -1.0954         12.0000
   20.0000          -1.4605         11.0000
   21.0000          -1.4605         11.0000
   22.0000          -0.3651         14.0000
   23.0000          -0.3651         14.0000
   24.0000           0.3651         16.0000
   25.0000          -1.4605         11.0000
   26.0000          -0.3651         14.0000
   27.0000           0.3651         16.0000
   28.0000           0.7302         17.0000
   29.0000           0.3651         16.0000
   30.0000          -0.7302         13.0000
   31.0000          -0.7302         13.0000
   32.0000           0.3651         16.0000
   33.0000          -0.7302         13.0000
   34.0000           0.3651         16.0000
   35.0000          -0.3651         14.0000
   36.0000          -2.1908          9.0000
   37.0000           1.0954         18.0000

  190.0000          -1.8257         10.0000
  191.0000          -1.4605         11.0000
  192.0000           0.3651         16.0000
  193.0000           1.0954         18.0000
  194.0000          -1.0954         12.0000
  195.0000           0.0000         15.0000
  196.0000           0.3651         16.0000
  197.0000          -1.4605         11.0000
  198.0000          -0.3651         14.0000
  199.0000          -1.0954         12.0000
  200.0000           1.8257         20.0000
```

Figure 2.1B (continued)

```
    MEAN              VARIANCE         SIGMA
     0.0037              0.9822          0.9911


 ****
 CUMULATIVE DISTRIBUTION
 DEVIATE             FREQUENCY             CUM FREQ            CUM PERCENT
 -5.4772              0                     0                   0
 -5.1121              0                     0                   0
 -4.7469              0                     0                   0
 -4.3818              0                     0                   0
 -4.0166              0                     0                   0
 -3.6515              0                     0                   0
 -3.2863              0                     0                   0
 -2.9212              0                     0                   0
 -2.556               1                     1                   5.00000000E-03
 -2.1909              4                     5                   2.50000000E-02
 -1.8257              2                     7                   3.50000000E-02
 -1.4606              14                    21                  .105
 -1.0954              17                    38                  .19
 -.7303               16                    54                  .27
 -.3651               32                    86                  .43
  0                   29                    115                 .575
  .3651               33                    148                 .74
  .7303               16                    164                 .82
  1.0954              11                    175                 .875
  1.4606              14                    189                 .945
  1.8257              6                     195                 .975
  2.1909              5                     200                 1
  2.556               0                     200                 1
  2.9212              0                     200                 1
  3.2863              0                     200                 1
  3.6515              0                     200                 1
  4.0166              0                     200                 1
  4.3818              0                     200                 1
  4.7469              0                     200                 1
  5.1121              0                     200                 1
  5.4772              0                     200                 1

END OF PROGRAM
```

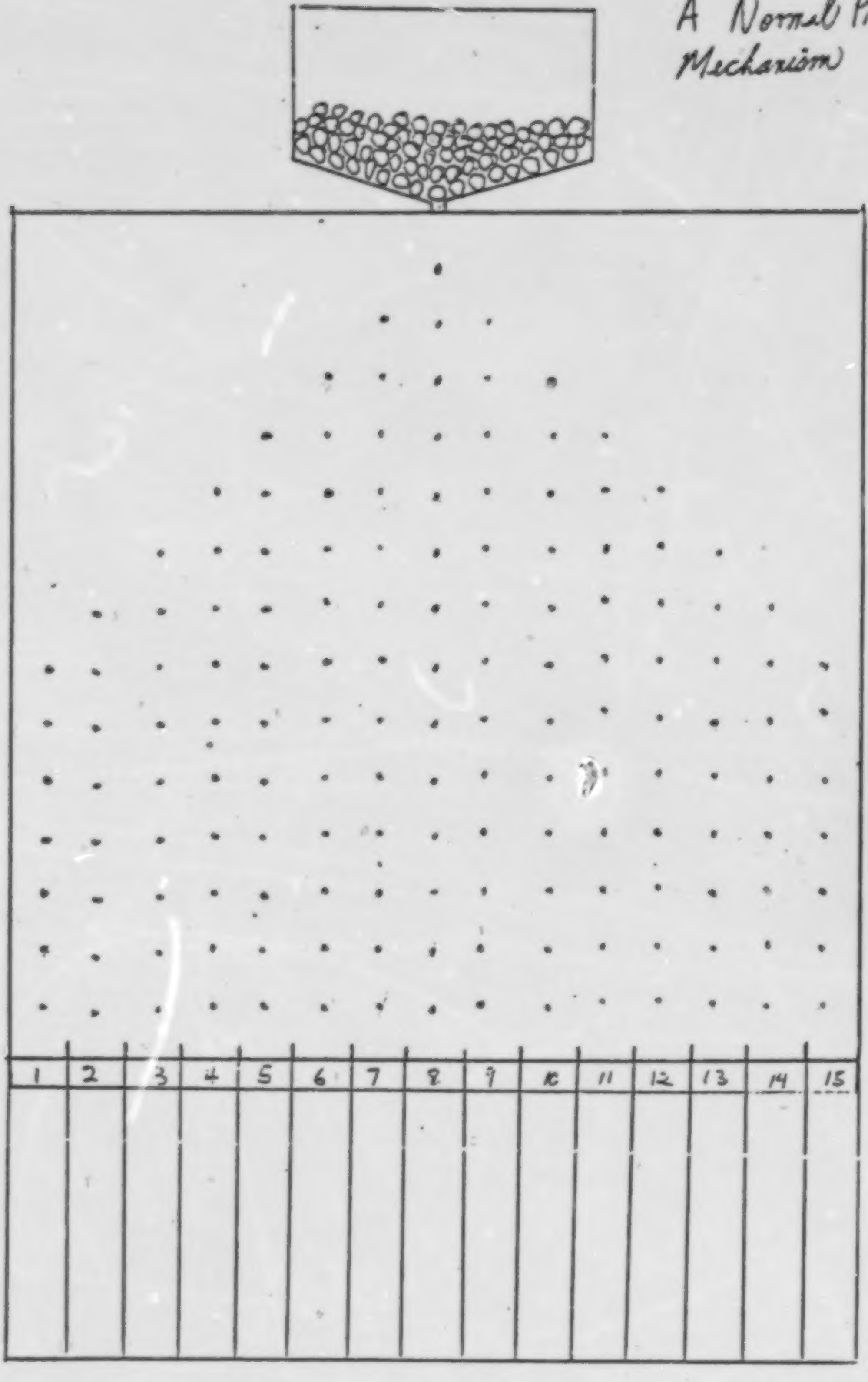
the sum of a large number of independent (unrelated) stochastic components, each generated by the same mechanism. That mechanism may in fact be any of several kinds. We shall now discuss a device in which the binomial, Bernoulli process mechanism is used repeatedly to get the sum.

Suppose we consider a number, say 30, observations on a Bernoulli process. The sum of the outcomes from those 30 observations will, over a large batch (say 200) such sums, be distributed according to the Gaussian or Normal distribution. The relevance of this idea to the study of education program impacts is direct. It should be apparent that the score of a student on an achievement test is itself a function of the sum of many (nearly) independent binary outcomes--items answered right or wrong. By a more powerful version of this same central limit theorem, it can be shown that those sums (test scores) will be distributed normally, as the number of terms of the sum increases.

A very clear and clever physical device displaying this idea is on exhibit at the Chicago Museum of Science and Industry. The exhibit resembles a vertical "pinball" machine. There is a glass window through which the audience looks into the device. An inch or so behind the glass is a wall, and there are a large number of steel nails stuck into the wall in a roughly square pattern, as shown in Figure 2.2. At the top of the window is a small opening through which rubber balls about the size of golf balls are dropped automatically, one at a time in rapid succession. As each ball falls, it hits one or another nail, and bounces either to the left or right. This is the Bernoulli process.

Figure 2.2

A Normal Process Mechanism


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

On its way down through the pattern of nails, each ball hits several nails, typically 10 to 15. As it hits each nail, the ball may flop either left or right. Thus, its final degree of leftness or rightness is the sum of these several independent events. At the bottom of the window, where the balls eventually come to rest, there is drawn on the glass an outline of the Gaussian curve. When 100 balls are dropped, they dependably spread themselves out so that they stack up closely in correspondence with the drawn curve. This regularity, in which the location of any single ball cannot be predicted, but in which the distribution formed by 100 sample balls can be predicted, is a vivid demonstration of what is meant by saying that we have probabilistic knowledge about the location of any single ball. For any single ball, we can state only that it is most likely to land somewhere near the middle, that its chance of landing far out to either side of the exhibit is small, and that it is equally likely to land on either side of the center.

In the real world phenomena studied by science, there are many situations where the outcome of a process can be seen as the result of the accumulation (sum) of many small causes. These outcomes thus can be assumed to follow the Gaussian distribution. Two frequently mentioned examples of such outcomes are personal characteristics such as height which are determined by the accumulation of a large number of genetic and environmental influences, and the errors of measurement made by astronomers or others in reading an instrument. Another example might be the scores on an achievement test containing a large number of items, each scored as right or wrong.

Let us now consider the parameters of the Gaussian distribution. The parameters are like the "tuning knobs" of the device. We could easily imagine a physical device with such tuning knobs as part of it. If we were investigating the operation of such a device, we would be quite interested in the settings of those knobs.

The Bernoulli processes model discussed earlier has, on each trial, only one tuning knob or parameter, which we had denoted as p. The Gaussian process model has, it turns out, two parameters. These parameters are the mean and the variance (or equivalently, the mean and the standard deviation). If we were in charge of a device for generating observations on variables which follow the Normal distribution, we would have to choose values for each of these two parameters in order to specify the data-generating structure. The parameter called the mean would determine the location of the outcomes along the entire continuum of real numbers. The parameter called the standard deviation would determine the amount of spread that our generated data points would have. Thus, any combination of settings of these two values would still generate data that follows the Normal model, even though the observed values might be quite different from set to set. Any particular specification of the parameters would define one structure within the family of Gaussian process models.

Thus far in our discussion of the Normal distribution we have spoken in terms of generating data given the parameters. This topic is, of course, the province of probability theory. As scientists, however, we are much more interested in the converse problem, namely that of gaining knowledge about the parameters of the process, given that we have observed some data. We are seeking to formulate specific structural models for a kind of phenomenon and to obtain knowledge about some of the parameters in those laws. Notice that whether the estimated values will remain constant from one situation to the next is an empirical question.

2.4 Conventional and Bayesian Logics of Inference

With all of this as an important broad foundation, we can now describe the differences between conventional statistical inference and Bayesian inference. To do this, we will make heavy use of what we shall call the "likelihood function". Our usage of this term will be somewhat different from its customary employment, but hopefully more consistent. The basic semantic distinction we wish to maintain is that when we are speaking of the chance of observing a particular data set given the values of the parameters, we shall use the word probability or probability density. On the other hand, when we are dealing with the description of knowledge about a parameter (or a parameter value) given a body of relevant evidence, we shall speak of the likelihood or likelihood density for the parameter.

The conventional usage of the term likelihood refers to only one of several possible situations where such knowledge is to be described. That is, among conventional statistical writers, the likelihood curve does describe the "plausibility" of various possible parameter values, given an actual, specific data set. It has also been used by Bayesians as the term to describe the same notion. However, the Bayesians, especially those inclined to use conjugate priors (of which more later), have not made the logically obvious extension to speak of the prior likelihood curve and the posterior likelihood curve. Instead, they speak of the prior and posterior probability. There are advantages to each choice of term. The word probability as a description of the belief distribution emphasizes the fact that the probability axioms and rules can be applied to these distributions. The word likelihood emphasizes the distinction between deductive inference (probability of various data) and inductive inference (likelihoods of various parameter values).

The need for a distinction on this point seems clear. R. Carnap for example, has coined the terms "probability$_1$" (inductive probability) and "probability$_2$" (direct probability) to distinguish the ideas. Novick and Jackson (1974) talk of "model densities" and "Bayes densities". Hacking has suggested chance and probability, which does have much appeal. In any case, in this report, and especially in the present chapter, we have decided that it is important enough to make the distinction between the deductive and the inductive modes of inference, and to draw attention to the similarity of the data-derived likelihood in the conventional analysis and the prior or posterior belief distribution of the Bayesian, that the word likelihood will be used for any inductive inference need. Note that there is also a distinct problem, already alluded to, lurking in the Bayesian literature--much of the work has been done within the context of discrete data-generating models, and for the comparison of discrete hypotheses. That is, Bayes's Theorem for continuous variables on both sides is rather seldom used in discussions of the logic of scientific inference methods. This appears to be largely an accidentally arising tradition, perhaps due to the historical prestige of laboratory experiments, but it has some consequences for the ease of becoming familiar with certain aspects of the continuous inference situations.

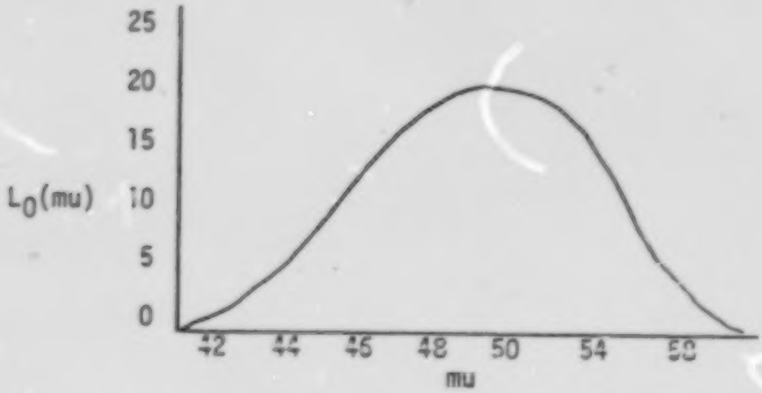
Perhaps one way to make the general point clearer is to take note that an expression like that given earlier for the binomial data probability is, when looked at entirely by itself, simply an expression with three symbols (S, N, and p), all of which may take on various values. Thus, we could say that this expression describes a _joint_ function of the data and of the parameters. When we hold constant the parameters and examine the conditional function for data values, we describe it as a probability function. When we hold constant the data and some of the parameters (_e.g._, S and N), then we describe this expression as a likelihood function for the remaining parameter (say, p). In our terminology, this will be called the _sample likelihood_ function. It describes the likelihood for the parameter, given the sample. That is, it relates the information in the sample data to the various possible values of the parameter of the data-generating model.

Where the Bayesians usually speak of a "prior", or a "prior probability", we shall speak of a prior likelihood. Finally, where the Bayesian speaks of the "posterior", or the "posterior probability", we shall instead refer to the posterior likelihood. The prior likelihood is a function, held by "YOU", which relates the information you have before analyzing the data under discussion, to your beliefs about the plausibility of the various possible parameter values in the situation under investigation. Similarly, the posterior likelihood relates the total information you have after analyzing the data under consideration (i.e., including all previous information or evidence and the information or evidence implicit in the data) to your beliefs about the plausible values of the parameter.

There are thus three distinct kinds of likelihood functions to be considered: (1) the prior likelihood; (2) the sample likelihood; and (3) the posterior likelihood. Each of them is described explicitly as taking various possible parameter values as arguments, and yielding numbers reflecting plausibility as function values. However, less obvious but quite important, any likelihood function also takes as input a particular set of information. This conditionality of probabilities on information is made clearer by the Renyi axioms (1955) than by the Kolmogorov axioms (cf. Novick and Jackson, 1974). The prior likelihood function takes as input information all the relevant evidence (which we may denote as I_0) available prior to the analysis of the present data. The sample likelihood function takes as input information just exactly the evidence provided by the present set of sample data (I_1). It takes this into account fully (within the framework of the formal generating model being used), but takes account of nothing else. Finally, the posterior likelihood function takes into account all the prior information (I_0) and all the sample information (I_1). Thus, the information input to the posterior likelihood function is $I_0 + I_1$. We denote this as I_2.

Once we recognize that these functions do have as input a set of information or evidence, it makes sense to think of them as being conditional on the state of information. In other words, the likelihood function has information and some process parameters as its parameters, and a process parameter as its argument. However, although the parameter values used as argument are continuous in most cases, the states of information are only of three kinds: prior, sample, and posterior. Other possible conceptual frameworks are of course readily imaginable. However, they are not needed here and will not be explored.

As indicated already, the likelihood function given any set of data can be thought of as a relation between two variables. Usually, both variables are continuous. Thus, the relation lends itself to graphic representation. For instance, consider the possible values of the parameter mu, the mean of a Normal data-generating process for, say, achievement scores among fifth graders. That likelihood curve, prior to analysis of the particular data set, might be graphed as follows:



The X-axis in this graph represents possible values of mu, the mean of the process. The Y-axis indicates the values of the likelihood function prior to the data analysis. Notice that the subscript 0 on the function symbol indicates the conditionality on the prior information. However, there is one more notion to be added. In the form in which they naturally arise, likelihood functions have varying metrics. That is, the total area under a likelihood function in one context might be quite different from the total area in another context. To avoid this lack of comparability, it is useful in some applications to rescale the likelihood function so that the total area under the curve for all possible values of the parameter is 1.0. For most purposes, the rescaling involves no loss in generality, and it does make the functions L_0, L_1, and L_2 directly comparable at any given point.

To make clear that this is sometimes done, we will often refer to the rescaled likelihoods as L_0^*, or L_1^*, or L_2^*, respectively. The asterisk makes clear that these are standardized conditional likelihood functions. A point which is relevant here is that, for the prior and posterior likelihoods, this rescaling to an area of 1.0 makes them legitimate entities under the axioms for mathematical probabilities. This fact thus allows us to use Bayes's Theorem to manipulate them. This in turn is the major reason why they are often described as probabilities. In short, L_0^* is interpretable as a prior probability distribution, and L_2^* is interpretable as a posterior probability distribution.

Interestingly, there are no usual examples of rescaled sample likelihoods. This is perhaps because the major development of this concept was done under the context of discrete hypotheses, where it is not reasonable to assume that the total area under consideration sums to all the area conceivable, and because the use of this function by classical statisticians has been mainly in terms of comparing points along it as ratios.

As has just been suggested, the fundamental difference between Bayesians and conventional statisticians, in terms of their philosophies of inference, is that the Bayesians customarily use L_0^* and L_2^* as well as the sample likelihood. The classicists reject both of these because they find the idea of treating a parameter as a random variable unsatisfactory. This may be because they have neglected the distinction between the ontological and epistemological usages of this notion, or it may be simply because of an obsolete view about the nature of physical science. Whatever the explanation, this is the fundamental distinction between them. The Bayesian is not so labeled because he believes in Bayes's Theorem and the sampling-theorists do not. Rather, it is because he has a different view of the nature of scientific inference and knowledge and consequently finds much broader application for this theorem. Indeed he makes it the cornerstone of his formal algorithms for inference. Novick and Jackson (1974), and L. J. Savage (1964) each make this point fairly clearly, but it is often not appreciated.

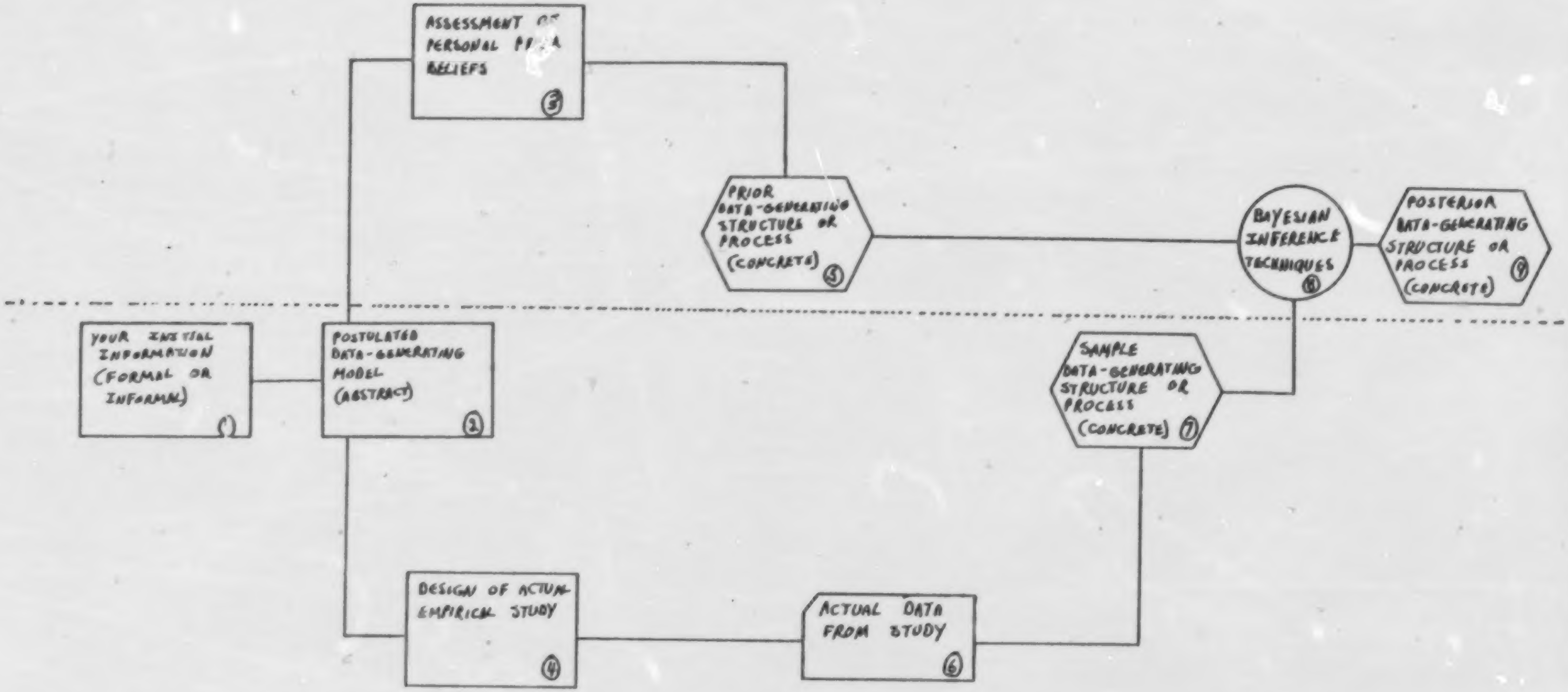
2.5 Scientific Research Methodology: Conventional and Bayesian

We turn next to a consideration of the steps involved in applying the Bayesian paradigm. Incidentally, it should be emphasized perhaps that it is a paradigm--in Thomas Kuhn's sense of being a fundamentally distinct vision and definitional approach to an area of investigation. Obviously, it is not a substantive paradigm, but rather a paradigm for scientific inference; it is a methodological and epistemological paradigm. Many of the steps in applying this Bayesian paradigm to a concrete research problem are parallel to those involved in the conventional statistical approach. Others, however, are not matched by any aspect of the usual statistical inference procedures. Not surprisingly, these distinctive procedural aspects of the Bayesian approach revolve mainly around the use and interpretation of the two likelihood functions which are not directly based on the present set of data. These are the prior standardized conditional likelihood, L_0^*, and the posterior standardized conditional likelihood, L_2^*. To simplify future reference to standardized conditional likelihood functions, hereafter we shall refer to them using the abbreviation "SCL".

The general flow diagram presented in Figure 2.3 gives an overview of the steps involved in using the Bayesian paradigm within a scientific research setting. We shall discuss it in some detail.

Note that the whole investigatory cycle begins with "your" state of initial information about the phenomenon, as denoted in unit 1 of Figure 2.3. The notational device of speaking of "you" is borrowed from that used by Howard Raiffa in his work Decision Analysis. Raiffa used this term to emphasize the personal and individual character of these prior beliefs about the phenomenon. This distinguishes him

Figure 2.3


YOUR INITIAL INFORMATION (FORMAL OR INFORMAL) ①
POSTULATED DATA-GENERATING MODEL (ABSTRACT) ②
ASSESSMENT OF PERSONAL PRIOR BELIEFS ③
DESIGN OF ACTUAL EMPIRICAL STUDY ④
PRIOR DATA-GENERATING STRUCTURE OR PROCESS (CONCRETE) ⑤
ACTUAL DATA FROM STUDY ⑥
SAMPLE DATA-GENERATING STRUCTURE OR PROCESS (CONCRETE) ⑦
BAYESIAN INFERENCE TECHNIQUES ⑧
POSTERIOR DATA-GENERATING STRUCTURE OR PROCESS (CONCRETE) ⑨
time →

somewhat from a slightly less personalized view held by many other students of the Bayesian idea, notably Harold Jeffreys and more recently Melvin Novick. The personalist vs. necessary debate is, from a philosophical standpoint, an important one. Its protagonists see themselves as arguing for individual personal judgment, subjective and guided only by one's own standards, versus judgments which in fact are either logically warranted or not warranted by the evidence available. This latter view, which is also very similar to Rudolf Carnap's ideas about degree of confirmation, is perhaps more suited to discussions carried on in a scientific context, where the individual judgment is typically validated by acceptance given by other specialists. We mention this distinction here because it does become relevant from time to time as the Bayesian approach is applied. For most purposes, however, we can simply avoid taking a clear stand on this issue. We view the judgments as actually made by one person, "you", but we are quite prepared to consider that they must be validated by the approval of others.

Either through such collective validation or through careful introspective analysis, the obligation placed on "you" is severe. In theological terms, "you" must "form a right conscience" about the phenomenon in general, and must "examine your conscience carefully" to see how you should view the situation under study. In some respects and in some circumstances, this is relatively easy; other times, it is quite difficult.

Occasionally, the initial information "you" possess about the phenomenon may be quite simple, coherent, and already formal. For instance, "you" might be concerned with coin flips, and you might know simply that the coin under study, when flipped 50 times, yielded 30 heads and 20 tails. This illustrates simple, coherent, formal information. Alternatively, you might know that this same coin, when flipped 50 times by Person A, yielded 40 heads and 10 tails, but when flipped 50 times by Person B, yielded 18 heads and 32 tails. Such information would still be simple and formal, but it would be much less coherent. Notice too that this initial information is called simple because it is expressed directly in terms of events similar in character to the outcomes which might be observed (*i.e.*, the data) in the study "you" expect to perform. In other words, this particular prior information is directly representable in terms of "*data-equivalents*."

In other situations,--unfortunately those more typical of the problems studied by social scientists--the initial information available is *not* simple, *not* coherent, and *not* formal. It arises instead from many disparate, diffuse, and informal bits of experience and communications, all occurring in different contexts and different formats. Any synthesis of this kind of information in "your" mind is necessarily vague, informal, and complex. Usually, such a synthesis is developed without full advertance or reflection, and there is no assertion of an equivalence between the prior information and any well-defined data set that might be imagined. Yet, whether given in terms of prior data-equivalents or simply in terms of a vaguely justified judgment, the Bayesian paradigm *requires* that one express his prior information in the formalized notation of a prior SCL (L_0^*), often referred to as a prior probability curve, or a prior belief density. This process of formalization (corresponding to unit 3 in Figure 2.3) is called assessment; it is often rather problematic in practical situations.

Logically coming even before this assessment, and shared by the conventional analyst and the Bayesian, is the work of specifying the formal data-generating model. This activity is shown as unit 2 of Figure 2.3, and is located below the dashed line to indicate that it is shared by both paradigms. Good discussions of this specification activity can be found in K. Land (1971), and in R. B. Braithwaite (1953). The data-generating model that results is an abstract set of relations among variables. The model will, given certain specified inputs, yield output values of the same type as found in the data of the study.

The activity of specifying such a model involves two kinds of tradeoffs. First, "you" must choose, simply on substantive grounds, a single formal representation of the immediate process of creating data observations that "you" believe is likely to be correct in the relevant (itself a word which hides many complex issues) ways. At the same time, and perhaps less obvious, "you" usually must also embed that proximate model in a broader, global model of the phenomenon, so that a sizeable theoretical structure in fact must be formalized. This latter activity--formalizing the global model--is typically not strictly necessary in the case of a true experimental study, where treatments are assigned at random to individuals. This may explain why the need for it in field studies is not more widely appreciated.

The attempts to perform this model specification, both the proximate and the more global aspects, sometimes make clear that the investigator ("you") has no definite conviction that one pattern of relation holds rather than another. However, neither the conventional nor the Bayesian paradigm can move further unless a commitment, however tentative, is made to one well-defined formal model. In other words, what may be truly a somewhat amorphous and disorganized set of impressions about the phenomenon must be transformed into a clear and coherent formal model.

Let us make this same point a slightly different way. The conventional statistical inference logic requires "you" to commit yourself to a particular form of the relationship, and also to a particular single value (hypothesized) of the parameter. It does not allow any intermediate degrees of belief about either the form or the parameter value. The Bayesian paradigm, on the other hand, does allow "you" to express degrees of belief about the values of the parameter. However, with respect to making a specific commitment to one formal model, the Bayesian and conventional paradigms are completely alike.

It should be mentioned that, considering broader research methodology, there are some ways to avoid being too constricted by this requirement, to preserve a fair amount of flexibility. One of these ways is to choose a formal model which involves many unknown "free" parameters. Thus, many of the points at which uncertainty about the phenomenon exists become represented by these parameters, and there is less regidity in the model. A second approach is to be very tentative in the degree of commitment to any single model, and to couple this tentative commitment with the idea of entertaining several distinct models, perhaps incompatible ones, rather than firmly postulating any single one. This latter approach, used by scientists for centuries, has been labeled the "method of multiple working hypotheses" by T. C. Chamberlin, and was canonized by John Platt as the ideal type of scientific approach (1964). In some respects, it comes very close to the logic of the Bayesian paradigm when there are a finite number of distinct hypotheses. Naturally, there are drawbacks to either of these attempts to avoid premature exclusion of model forms. Basically, they each place a much greater burden on the data to provide distinguishing evidence. A model in which there are many free parameters is often deprecated by researchers as being too hard to reject. It can, unfortunately, be adapted so as to be consistent with a wide variety of patterns in the data, and thus may serve less as a means of simplifying than as a means of relabeling the phenomenon. Similarly, the method of multiple working hypotheses, though somewhat tighter, also requires a great deal of data (in quantity and quality) if the hypotheses themselves are at all complicated. In short, with either version of the method, the study of complex systems outside of laboratory experiments poses major difficulties.

This is perhaps one reason why the social sciences have lately begun to founder a bit. Their attention has shifted from relatively simple questions of bivariate relationships to a concern with the behavior of entire systems of inter-locked variables (cf, Cooley and Lohnes, 1976, for an illustration). This shift in interest has corresponded to a shift in methods from univocal analysis like that of the laboratory experiment to complex, multivariate, analysis more close to that considered in systems theory. This idea of what may be described as "systems nested within systems", in which the micro-level variables depend stochastically upon certain parameters, and these parameters in turn are regarded as variables (perhaps stochastically determined) in a model pitched at a more macro level, is quite important. It highlights one key issue on which the paradigm of Bayesian inference is incidentally helpful. That is, the utility of the Bayesian paradigm in this respect is not so much an intrinsic part of it, but simply a latent consequence of the way in which it leads to viewing the world. As is apparent, it links very closely to issues involved in choosing appropriate levels (perhaps several) of aggregation. But, because the Bayesian approach is still quite foreign to most social researchers, these workers tend to find that they have no really satisfactory way of dealing (at present) with many of the questions that they have set for themselves. We simply lack the instrumentation or manipulative control to carry out studies which either would collect enough data (again, in quantity and quality), or would make simplified analysis applicable by simplifying the phenomenon actually studied via laboratory-like control. Perhaps worse yet, our prevailing paradigm of inference tends to cause us not to even perceive this quandary fully.

That limitation is obviously a matter of degree, and it will be continually challenged by researchers in a variety of areas, through creative developments in instrumentation and design, and through the gradual accumulation of data bases and data collection routines. However, its existence must be recognized and emphasized, as a guard against the tendency to tilt at windmills.

Perhaps an equally important general point to be made about the model specification activity is that unavoidably some flexibility of understanding is lost in the process of choosing a specification. Some other formal models which might have been developed and entertained on the same vague initial information are instead simply dropped from further consideration. The scientist's mind turns to other topics. This is a mixed blessing. By narrowing down the alternatives, even to the point of considering only a finite set of continuous parameters, one loses potential richness of understanding, but gains potential clarity, communicability, and empirical testability.

All of this has concerned the strictly substantive decisions and tradeoffs. There is a second broad class of tradeoff also to be considered. In general, this is the choice between seeking complete representation of the phenomenon versus seeking manageability and communicability of the model. The manageability depends on the intellectual apparatus (general mathematics, statistical theorems, logical proofs, etc.) that is available. The communicability in turn depends on the representational aids that can be used. Pictorial approaches, and in particular the possibilities of computer dynamic graphics, may make possible a considerable increase in what can be achieved, but at a high cost in dollars.

The decisions to be made on this dimension of coverage versus manageability can be seen as decisions as to what may be omitted, or treated in a much oversimplified way. These decisions can focus on (1) certain aspects of the real phenomenon, and how they should best be dealt with in the modeling; and (2) certain technical and methodological concerns which are not isomorphic with the actual phenomena, but which will affect the interpretations drawn from the behavior of the model. These decisions are difficult. There is always a need for including more material than the analytic resources will support. Moreover, the decisions are made on the basis of day-to-day increments in the relevant evidence. Information arriving during a project, but after the basic model and design have been established, may cause "you" to feel the decisions should be modified. Yet, once an investigation has gotten underway, some kinds of changes cannot be made without enormous cost. There is little general advice that can be given; this is a judgment that any research director makes in the concrete situation. Yet, perhaps a neglected point is that the primary role of a project supervisor, or of an expert consultant on a larger research study, may be that of providing a combination of critical suggestions, and of personal support to the researcher confronted with uncertainty. In the former role, the consultant represents the community and its intellectual interest in the project. In the latter role, he again represents the community, but this time with respect to its supportiveness for the researcher's best efforts.

One broad consequence of this entire discussion is that two sets of notions--the activities involved in model specification, and the detailed content of the resulting model, need to be made more explicit. Without that explicitness, both as to content and as to the actual evolution of the model, the readers will be unable to appreciate the worth of the model, and its limitations, properly. A change in reporting of this sort might be somewhat threatening to many researchers. Also, it would be inconsistent with present scientific report style, in which the hypothesis is produced, like a proof in an advanced mathematics text, with a bare minimum of clues as to its origin or implication. But that extremely terse style of presentation is appropriate only when there truly is a definite logical chain underlying the development, so that the steps themselves are not central. In most social science research where complex models are specified and used, the development of those models is governed not by pure logic, but by what Herbert Simon and others have called "bounded rationality". Thus, to appreciate the resulting model fully, it is important also to be aware of the origins and aspirations of the developer. More specifically, what is needed is a clearer discussion of the alternative formulations that were briefly considered, the sequence of stages for considering various aspects of the model, the major pros and cons on each representational choice that was made, and--perhaps most important--a statement of the developer's ("your") confidence in each particular aspect of the model.

In fact, of course, many such choices are nearly arbitrary. They are made strictly for reasons of computational convenience, or perhaps because of an idiosyncratic experience which was dominant in the memory of the developer. To admit such behavior openly given the present norms of scientific practice would be gauche, and more important, would damage one's career. However, this is, it appears, the way all of us do behave when confronted with the need to deal with an unavoidably complex analytic challenge, and so the communal norms on reporting must change if we are to achieve self-awareness collectively.

As Figure 2.3 makes clear, this step of specification of a model is one that must be performed by conventional as well as Bayesian analysts. The attempt to use the Bayesian approach has drawn more attention to it, but that is purely a coincidence. Any improvements in practices related to model specification will benefit equally the adherents of each paradigm.

The next step described in Figure 2.3, the assessment of the prior SCL for each focal parameter of the model, is unique to the Bayesian paradigm and has already been touched upon briefly. We shall use the word assessment strictly to refer to a process of converting personal, perhaps judgmental, information about a phenomenon into an actual likelihood distribution over all possible values of each focal parameter. For the sake of simplicity, we shall usually speak only of a single such parameter, because a number of complications arise when several parameters are jointly in need of assessment.

Assessment may be approached in a variety of ways. Two in particular deserve mention: (1) the method of subjective indifference between alternative bets or gambles (drawn mainly from decision theory); and (2) the method of equivalent prior samples. The latter approach, somewhat less well developed, is nevertheless the one which appears most suitable for the application of the Bayesian paradigm to scientific questions. This latter approach is the one favored by Jeffreys, Keynes, Carnap, Novick, and apparently by Box and Tiao (1973) also, though it is not the usual treatment described in books on Bayesian decision analysis. Some actual computer packages designed to assist with and to structure the assessment procedure (such as the MANECON package or the CADA package) make use of both approaches simultaneously. They deliberately elicit redundant assessment information as a means of checking on the consistency of the assessor's assessments.

The equivalent prior samples approach (i.e., representing knowledge in terms of data-equivalents) amounts to confining oneself to the class of prior SCLs known as the conjugate priors. Such priors are, by definition, distribution curves whose form can be represented by assuming that the form is the same as the form of the data-generating process, and by choosing parameter values for these priors which happen to correspond perfectly with particular patterns of sample information. Using such a class of priors greatly simplifies the mathematics of Bayesian analysis, since there is no need to do numerical integration. The combining of prior SCL with sample SCL is achieved through some simple analytic manipulations, and, better still, the resulting distributional form (the posterior SCL) is also of the same form as the prior SCL, and so the entire process can be repeated. This makes it quite simple to cumulate the results of several successive studies.

The use of conjugate distributions seems more reasonable in scientific analysis using Bayesian inference, since the usual situation within science is that there is prior data available on the issue. Usually that data does not bear directly on the present study, and so the implications of it must be modified before being taken as formally equivalent to the investigator's ("your") prior beliefs. However, the basic principle that prior knowledge is primarily empirical knowledge is one which strongly argues for the use of conjugate priors (perhaps somewhat modified from what might seem the most obvious formalization) in scientific uses of Bayesian inference. This idea--the development of modified conjugate priors--is one which has been mentioned in conversation by some researchers, but which has not been studied in any detail. Thus, one topic worthy of future investigation is just exactly what impact is exerted on a later study in an area by the data gathered in earlier studies. Is there, for example, a decay rate, so that old data is of little relevance? Is there some implicit scale of similarity of situation which determines the degree to which existing data is "discounted" before being taken into consideration?

Another area for investigation is the feasibility of using the more general, non-conjugate prior distributions as representatives of the prior state of information. This option increases the flexibility available to the Bayesian analyst, since with it any pattern of prior beliefs can in principle be handled. However, a number of technical questions have been neglected by the casual statement that the Bayesian approach can handle any prior distribution whatever through using numerical integration techniques. First, it is not known what degree of detail is feasible when an assessor is making the actual assessment and is using an arbitrary prior SCL. A second concern is whether asking for a report of a density curve is preferable to asking for a cumulative curve. There has been some work on these areas, but actually very little has been done with the aim of applying the Bayesian paradigm to scientific analysis (cf. Winkler, 1967; Hogarth, 1975; Hampton et al., 1973).

Such studies of assessment are quite feasible at present, as is evidenced by the fact that the MANECON package of programs available from the Harvard Business School utilizes the idea of an arbitrary prior distribution as its basic approach to assessment. At present, this research task would be a sizeable one to plan and perform. However, imminent developments in computer graphics technology may be quite helpful here. The use of a light pen actually to draw a curve, or one of the now available devices for conveniently inputting a curve into a digital analyzer mechanism, might be the kind of technical aid which would greatly facilitate the assessment process. In fact, in the past it has often been just such developments which have transformed ideas previously thought to be interesting but impractical into the mainstay of a later style of work.

A third topic within the general area of assessment of priors is the possible substitutability of the best member of a conjugate family for an initially assessed arbitrary prior. Here too the actual investigation of this question is presently feasible, and the MANECON package plus the Dartmouth Time-Sharing System program DAMAXS provide a basic computer technology for such an investigation. Again, however, there would need to be considerable effort devoted to this as part of a substantial and several-year-long research program.

The next step in the logic of a research project again is parallel in the conventional paradigm and in the Bayesian paradigm. It consists of designing an actual empirical study--choosing the configuration of data that would be obtained, and then as nearly as possible arranging to obtain it. It is shown as unit 4 in Figure 2.3. For both the Bayesian and the conventional statistician/scientist, this is a complex, delicate, and uncertain endeavor. It requires a blend of several talents, and the commitment of much hard work. The design considerations are those of tradeoffs between feasibility of procedure or access, substantive plausibility of certain assumptions, potential clarity of evidence, administrative and personnel requirements, and time-cost limitations. In most ways, the conventional and the Bayesian paradigms lead to similar choices, since both are seeking the best possible data. However, they may lead to a difference in the investigator's ("your") attitude on one important topic.

The Bayesian paradigm, because it conceptualizes the parameters as themselves being variables, will encourage the research designer to think more in terms of what might be called "contingent relationships". In the terminology of analysis of variance, these might be described as interactive effects; in medical contexts, we might refer to potentiation; in educational research, one instance of this kind of thinking has been the idea of "aptitude-treatment interaction". In other words, for no formal reason, but just as an incidental consequence of his view of knowledge and of the kind of phenomena under study, the Bayesian analyst tends to be more comfortable in dealing with circumstances which cause the parameter values to vary. He, compared to the conventional statistician, is more prepared to interpret fluctuations in parameter values as real. He may regard them as occurring systematically, and view the parameters (at the level he is focusing on) as simply being variables in a larger-scale conceptualization of the general phenomenon, or he may view those parameters as fluctuating somewhat randomly between situations. Perhaps a concrete, though nonsocial, example here would be helpful. When the Environmental Protection Agency measures and reports on the gasoline mileage of all models of new cars, the report can be described as being expressed by two parameters for each type of vehicle, the city mileage and the highway mileage. Yet, most users of these statistics fully appreciate that there are systematic differences in the value of these parameters between say, Cadillac and Volkswagens. They also appreciate that the parameter values probably fluctuate somewhat between individual samples of a Cadillac, and between individual samples of a Volkswagen. In short, two Volkswagens constitute two instances of a highly complex system. There is little reason to expect exact identity of performance parameters in the two systems, although there is strong reason to expect them to be relatively close. A Bayesian statistician would find this logic entirely familiar, and would be more inclined to use it than would a conventional statistical analyst. The conventional statistician might retort that he handles the same idea by speaking of the "components-of-variance" (Model II) version of analysis of variance, and this would be logically quite correct. In practice, however, the Bayesian is more open to this kind of thinking.

There have been some instances in sociological research where the researchers have sought to treat parameters of a process at one level as variables considered at a different level. An early and very impressive example of this idea is James Coleman's insight, published in his 1964 book on mathematical sociology, that we should regard many kinds of social phenomena as stochastic processes. Such processes (of which the Bernoulli process discussed earlier is one very simple example) are completely described once we have given a numerical value for all necessary parameters. However, and this is Coleman's basic insight, such processes are explained only when the fluctuation in those parameters across situations can be predicted from a formal model, either deterministically or stochastically. Thus, for example, Spilerman's 1971 paper on the occurrence of civil disturbances followed the logic of the Coleman approach, and provided a stochastic explanation for fluctuation in the process parameter.

In educational research, this general idea might take the form of viewing schools as the systems of interest, each characterized by some parameters of the production process for achievement. Fluctuation in these parameters across schools then could be seen as the relevant way of conceptualizing the nature of the process. Crude attempts to employ this kind of reasoning were made in two earlier studies in which I have been involved (Fennessey, 1970; Heyns, 1971). In both of these studies there were between 30 and 50 schools in the dataset, and our interest was focused on whether certain processes operated the same way in each school. Operationally, this was framed as the question whether the regression slopes in our specification of the proximate model were the same in each school. Without going into detail here, we can simply summarize by saying that in both studies, fluctuations in slopes were found, but these fluctuations appeared to be random (unconnected with any property we could identify). Although we were personally prepared to draw the conclusion that the fluctuation was in large part stochastic and non-systematic, our reliance on a conventional paradigm of inference in both studies made the discussion unorthodox, defensive, and awkward. It also prevented us from capitalizing on the point made more clearly by Spilerman (who was dealing with a simpler but more elegant model), that the concentration on explaining fluctuation in the process parameters makes it somewhat irrelevant how much variation is accounted for by the proximate model, or in other words, what the magnitude of the disturbance variable in the proximate model turns out to be.

As we have said, our way of dealing with these issues while using the conventional paradigm of inference was awkward. However, if a Bayesian paradigm had been used, this approach would have been not only reasonable and natural, but even obvious. This is not an inherent feature of the Bayesian inference logic, but rather something we believe will be a de facto correlated feature of it when it is applied to the scientific analysis of complex empirical systems. Moreover, this kind of investigation, in which there is observed a sizeable fluctuation in the parameters of a relationship, has become (especially in the last two decades) important in almost all fields of behavioral science research specialization (e.g., school effects studies; studies of organization dynamics in industry; studies of status attainment and labor force processes; social psychological studies of attitudes, etc.).

The remaining stages in the research cycle, whether in the conventional or in the Bayesian paradigm, are more formalized and codified than the preceding ones. The Bayesian combines his formal model with the assessment to produce a prior structure (Unit 5 in Figure 2.3). In both paradigms, the data is actually collected and summarized (Unit 6 of Figure 2.3). An analysis is carried out, and yields sufficient statistics with respect to the postulated form of the data generating process (Unit 7 of Figure 2.3). The conventional statistical analysis is essentially completed with the calculation of these statistics. Further interpretative transformations of them may be done, but the evidence used by the conventional approach in making its conclusions is, formally, just that from the data collected in the same study.

The conventional analyst will often carry his thinking beyond the formally derived numbers, and offer some informal conclusions or reflections as to how the results from the study at hand relate to earlier work on the same or similar topics. Such musings are helpful, but the key feature of them we wish to emphasize is that they are informal and thus somewhat imprecise in meaning.

The Bayesian analyst, in contrast, carries out a corresponding analysis on his data to obtain the sufficient statistics. He then, however, uses these statistics to derive the sample likelihood, and as the climax of the operation, feeds this sample likelihood and the prior SCL together into his Bayes's Theorem machinery (Unit 8 in Figure 2.3). That machinery then cranks out, in a reproducible and defensible way, the posterior SCL (Unit 9 in Figure 2.3). Thus, the Bayesian obtains as his result something which formally and correctly merges the investigator's prior beliefs with the implications of his data, and which presents those implications in all their richness. The posterior SCL is an entire distribution, not a single number. Thus, it is complete, and pictorially clear, but somewhat awkward to deal with as a totality.

One interesting feature of the Bayesian approach--though again one which is more a matter of emphasis than of actual difference between it and the conventional school--is that the Bayesian approach can sometimes lead to a situation in which it is clear that in a sense one knows the value of the parameter less well after the study than before. This would be the case, for example, if the observed data indicated that the parameter was most likely to lie in a region quite far from the regions favored by the prior SCL. The same actual situation can and does arise among conventional analysis workers, but is often not perceived clearly or correctly. To capture this kind of situation rather than letting it slip through the cracks of the formal techniques is one incidental advantage of a Bayesian paradigm.

3.0 Implementing Bayesian Analysis Procedures

3.1 Modeling the Academic Achievement Process

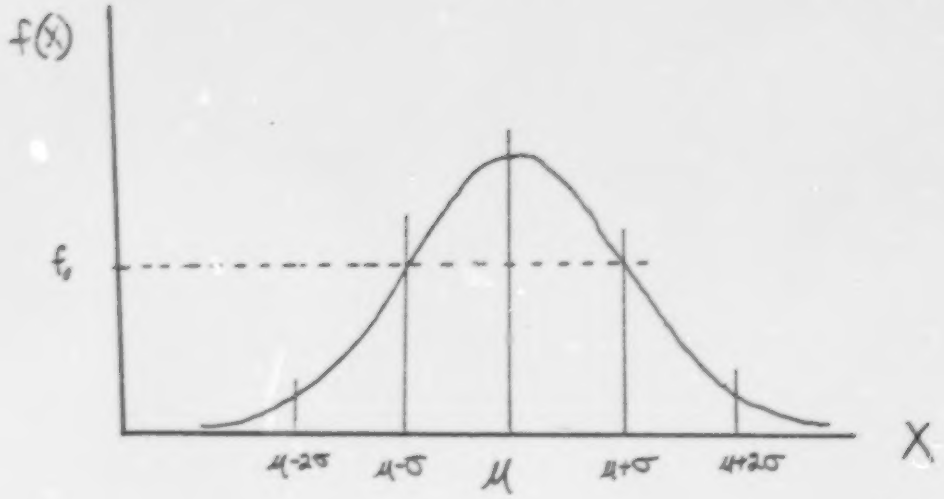
As has already been suggested, the substantive focus of interest in this research project was on the kinds of impacts exerted by various programs which seek to raise academic achievement of public school students in the elementary grades. To pursue this investigation, the first important task necessary was to specify a formal data-generating model for the process by which achievement scores are generated. This specification activity is complex, with aspects that are often neglected. Typically, a researcher working within the conventional traditions of statistical inference simply notes that the outcome variable of interest, achievement test scores, is a conceptually continuous variable, with a roughly bell-shaped, symmetric distribution. Building on this, he immediately goes ahead to apply the logic of t-tests or of analysis of variance for the analysis.

This quick leap from concern with the substantive data-generating process to adoption of a statistical technique can conceal some problems. Working within the Bayesian framework, where techniques are somewhat less developed, we were correspondingly more slow to choose a particular model. Two kinds of complications became evident as a result. First, it became clear that the choice of score format used to index the achievement level can affect the patterns found in the data, and that there is no consensus among researchers as to the proper (substantively most appropriate) mode of scoring the tests. Second, the level of aggregation used to make the comparisons also emerged as a problemmatic issue. Again, this was one which has been largely neglected, even by thoughtful researchers, and consequently there is little available wisdom here either.

Thus, our choice of model was not without some difficulty. However, it did seem clear that the general category of models needed would be those which postulate the data as being generated by a process which can be modeled by the Normal distribution model. This would be true regardless of the more detailed choices as to the particular score format, and as to the level of aggregation most appropriate.

As mentioned earlier, the Normal distribution family of models incorporates two distinct parameters. One is a location parameter, usually the mean of the process. The other is a spread parameter, measuring the dispersion of the scores along the range of possible values. There are several forms possible for this spread parameter, notably the standard deviation, the variance, and the precision (defined as the reciprocal of the variance). Of these, the standard deviation is attractive because it is a "scale" parameter. That is, doubling the standard deviation will double the range of scores between any two density values. This can be shown graphically, as appears in Figure 3.1. In that figure, the dashed line indicates two points at which the height of the curved line (the density value) is the same. If the value of the parameter sigma were doubled, then the distribution again plotted on the same scales as in the Figure, the new curve would be identical to the old. Thus, these

Figure 3.1



two parameters mu and sigma are a good choice. They are usually referred to by the Greek letters μ and σ. We shall use that notation or the actual names mu and sigma in this discussion.

However, our work is not finished when we decide that the basic process which generates academic achievement scores can be modeled by the Normal distribution. We must then move on to the next question, which is that of comparing the outcomes of two such processes, to see in what ways their parameters appear to differ. This brings us to the first point at which we encountered a series of gaps in the available literature.

3.2 Developing an Analysis Strategy and Procedure

Partly because of disagreements about the fundamental issues involved in attempting to formalize ignorance, there is a fair amount of disagreement as to the proper choice of "conjugate" distributions for the Normal distribution in which both the mean, mu, and the standard deviation, sigma, are not known exactly. If an estimate of the standard deviation is available, then the conditional distribution of the mean, mu, follows the formula of Student's t. Likewise, there is considerable agreement in the general direction of using a prior distribution known as the inverse chi to represent the prior beliefs about the standard deviation, sigma. Thus, our choice, made partly on the basis of availability of technical discussions, was to use a set of priors which can be called the Student's t/inverse chi, and to link those to the Normal distribution as the data generating process.

What has not yet been discussed, however, is the question of representing the comparison between two distinct outcome distributions, such as might be expected in a study where one group of students is exposed to a new educational treatment, and another group is exposed to the conventional treatment (i.e., is used as a control group). By making the additional assumption that the impact of the treatment will be to raise the score of each child by the same number of points, we can see that this leads to a form of analysis in which we are doing a comparison of two means from two Normal distributions. Notice that this assumption that the treatment impact consists of adding a constant number to each student's score is a somewhat strong one, but is in fact universally made in the conventional modes of analysis as well. Thus, although we are not entirely satisfied that this formulation fully or most appropriately represents our speculations about the treatment impact, we have decided to use it. We do so in the knowledge that, though imperfect, it is perfectly parallel in its imperfections to the mode of analysis involved in conventional analysis of variance of such educational impact research data.

Operationally, therefore, our task is that of carrying out the comparison of two means, one from a treatment group, and one from a control group. This corresponds to the familiar Student's *t* test of conventional statistics, and thus is a fairly simple mode of analysis. No esoteric manipulations are involved in the *t* test as it is conventionally used, and there are none in our Bayesian approach either.

The Student's *t* test, however, comes in several varieties.

First, there is the simple situation in which the two groups constitute independent samples. This leads to a single t value. Second, there is the situation in which an observed mean is being compared with a theoretically specified mean. This again leads to a single value of t. Third, there is the situation in which a difference in means is being compared, where the observations are not independent; usually this is known as the paired-samples t test. Finally, there is a situation in which the investigator does not feel comfortable in assuming that the standard deviations in the two samples are the same. Under the assumptions that those standard deviations may be different, the strictly appropriate distribution for the analysis is not the Student's t distribution at all, but instead a distribution that has become known as the Behrens-Fisher distribution. However, two features of this latter distribution are of relevance. First, as the size of the samples increases, this distribution comes to resemble the Student's t distribution more and more closely. Second, there are available some approximating formulas which improve the correspondence between the Behrens-Fisher distribution and the Student's t distribution, by modifying the parameters used for the Student's t distribution. In short, even in situations where it is mathematically not the relevant form of the conjugate prior distribution, the Student's t distribution is handy as a useful approximation.

In view of the fact that there has been considerable discussion in the educational testing literature about the relationship between the mean scores and the size of the standard deviations across various populations (cf. Fennessey, 1974), we felt it necessary to be able to carry out comparisons between the treatment groups and the control groups which did not require the postulation of equal standard deviations in the two groups. This meant, in short, that we needed an analysis package which would provide Student's t and Behrens-Fisher calculations. At this point, we began to encounter difficulties.

Our review of the generally available statistical literature, and our searching of various major libraries of computer programs (including the IBM Scientific Subroutine Package, the Digital Equipment Corporation DECUS library, the Dartmouth Time-Sharing Service library, and several journals and monographs on applied numerical analysis) all indicated that only two packages of Bayesian analysis programs are presently available. Incidentally, this was the case in late 1974 when the project began, and is still the case now, in late 1976, with the exception of the work we have done as part of this project.

The two available packages are the MANECON system, developed at the Harvard Business School, and the CADA system, developed at the University of Iowa. The MANECON system has been available since about 1970, and has evolved slowly since then. The CADA system is more recent, becoming available in a first version in 1974, and still evolving rapidly. At present, a 1976 version has been developed which incorporates some enhancements worthy of note, but that version has not yet become available in a form suited to our equipment at Hopkins, and so has not been used in this project.

A few words about these two packages are in order, to indicate why neither of them was well-suited to our needs as research data analysts interested in the comparison of means from Normal distributions. The two packages are far apart in the philosophy and style they represent. The CADA package is highly tutorial in style. It has been written in the language BASIC, and uses relatively crude algorithms for many of the calculations. Most of the numerical integrations are done by Gaussian quadrature, typically ones using 16 points. The package is interactive, which is a major strength, but it is also quite verbose, and the existing versions do not provide for any abbreviation of the dialogue. Also, the package is stronger on the analysis of discrete models (binomial, multinomial) than on the continuous models such as the Normal. It does, however, include routines for dealing with the Normal distribution, the Student's t distribution, and the inverse chi distribution, thus making it generally quite helpful. One major drawback to the package in the versions we have seen, however, is that there is no provision for obtaining graphic output of the curve being evaluated. Also, in the versions we have examined, the CADA package does not offer any ways to assess an arbitrary prior, or to combine that prior with a sample likelihood via numerical integration. The version just now becoming available promises to remedy some of these deficiencies. In particular, there will be inclusion of a crude analyzer for the Behrens-Fisher situation, and some improved assessments for arbitrary prior distributions. Also, there will be the introduction of a utility assessment aspect, something which is irrelevant to the specific task of this project, but which does enhance the CADA package as a general analytic resource. However, the package still has some serious limitations for the kinds of application we have just been describing -- the kinds involved in the comparison of educational program impacts. Specifically; it still lacks a graphing routine; the Behrens-Fisher portion is crude; and the dialogue is so extended as to make it tediously slow to use it for research purposes.

In short, the CADA package is quite an achievement, but is aimed at meeting the needs of pedagogy more than of research, and of discrete model analysis rather than continuous model analysis.

The MANECON package from Harvard embodies quite a different philosophy. At the risk of using stereotypes, one might almost say that the differences in style between CADA and MANECON are those between an educator and a management scientist. The MANECON package was written in FORTRAN, and is operated at Harvard on a DECsystem-10, the same model computer available to the present investigator during the latter part of the present project. The MANECON package embodies the most careful and elegant algorithms available for each calculation it performs, and employs a very terse, almost cryptic style of dialogue and output. Although it does provide sketch graphs on request, one of its drawbacks is that it provides no convenient means of obtaining well-labeled output, nor of obtaining results on the line printer. In short, like the CADA package, the MANECON routines are designed to be interactive, which is desirable, but they lack any of the conveniences which would make them more suitable for use in a research project. Also, the documentation of the MANECON package, though extensive, suffers from two drawbacks. First, it was written using a terminology peculiar to Robert Schlaifer's approach, and was intended primarily for use by students at the Harvard Business School. Second, the documentation takes the form of a published book, which was printed in 1970, and by now has become slightly outmoded with respect to the capabilities of the current version of the programs. A final drawback of this package for our purposes is that there is no provision in it for dealing with a Behrens-Fisher distribution.

3.3 Developing a Prototype Analysis Package

The review of available packages just described, concentrating on the CADA and MANECON packages, was in fact a much more extensive task than is suggested by the brief report. Both of these packages are fairly complex, they are evolving, and they have not been widely used outside of their home institutions. Moreover, much of the mathematics on which they are based is quite unfamiliar to the average social science data analyst. Thus, the combination of technical problems in adapting programs to other machines, and in becoming familiar with their capabilities, plus the further work involved in "getting inside" each of them to ascertain the prospects for making some modifications which would improve their usefulness, led to the devotion of considerable effort to this phase of the work.

The overall result of this investigation was a decision that it would be simpler, and cheaper in the computing costs of executing the analysis, if we developed a new set of programs to perform the analysis we needed. In this development, we tried to draw where possible on the strengths of the two available packages, but to create programs which would be more useful for researchers. We believe such programs will have some general utility, even beyond the study of educational impacts.

Data analysts, whether working in the sampling-theory tradition of inference or the Bayesian tradition, frequently are called upon to interpret statistical results derived from a Normal distribution data-generating process. These situations arise usually so that the mean, mu, is the parameter of main interest, and the standard deviation of the process is not known exactly. Under these conditions, the sampling-theory analyst uses a Student's _t_ distribution, and he regards that distribution as describing the pattern of sample means to be expected over repeated (imagined) samplings in which it is assumed that the population mean is the postulated value. That postulated value usually is either an a priori choice, or is the sample mean. A Bayesian analyst, in contrast, also uses the Student's _t_ distribution, but he interprets it quite differently. The Bayesian regards this distribution as describing the probability of occurrence of different possible values of the process parameter mu, taking into account the evidence from the sample and whatever prior beliefs may be available. Notice that the important difference here is in the interpretation of the parameter. In the sampling-theory approach, the parameter mu is regarded as a non-stochastic quantity, but in the Bayesian approach it is seen as (at least as far as our knowledge about it is concerned) a stochastic variable.

Data analysis problems involving the Normal distribution as the data-generating process, and its parameters as the focus of interest, are common. For example, in addition to its usefulness in comparing means in two populations, the Normal process, and the resulting Student's t distribution of the unknown parameter, also appears in the interpretation of the regression coefficients derived via a least squares analysis of a linear model, if the disturbances are assumed Normally distributed.

Related situations occur in which the analyst compares two separate means, each based on data from a Normal distribution. If the two unknown process standard deviations can be assumed equal, then the Student's t distribution again is the proper basis for interpretation. However, if the two means are being compared in a situation where the two process standard deviations are unknown and not assumed equal, then the analyst should refer to the Behrens-Fisher distribution, as already outlined.

At present, however, analysts in either the sampling-theory or the Bayesian tradition who wish to show the detailed implications of their data find themselves somewhat handicapped. Although many published tables of Student's t distribution exist, those which are most easily available are suited only for reporting statistical significance tests or confidence intervals for one of the conventionally used p-values. That is, most published tables provide only the critical values of t for several tail percentage points (e.g., 5%, 2.5%, 1% and 0.1%).

Such tables have other limitations as well. They provide the critical values only for a "standardized" version of Student's _t_ (one which has a mean of zero and a scale factor of unity). Thus, the analyst must perform an extra step of calculation when interpreting his data. Also, most tables present exact _t_ values only for a few degrees of freedom. In some of the larger compilations (e.g., Pearson and Hartley, 1966, Table 9), cumulative probabilities are given for a large range of _t_ values, and for a large range of degrees of freedom. However, tables of the _density_ values for Student's _t_ are quite scarce.

In large part, limitations such as those mentioned reflect the nature of statistical tables in general. Obviously tables for every possible combination of potentially interesting values of the statistics would be prohibitively expensive to compute, and unwieldy to use. These limitations are not serious for an analyst who seeks only to compute a significance test, but they do inhibit a researcher who wishes to provide his readers with a richer and more complete appreciation of the meaning of his findings. The limitations, in other words, are particularly serious for analysts working within the Bayesian framework of inference; for them the whole analytic approach is premised on the idea that the entire posterior distribution of a parameter is the only proper and adequate representation of knowledge about it. Because of this belief, the Bayesian analyst prefers to show his readers a graphic yet detailed representation of the posterior distribution of interest.

At present, however, few technological aids are available to carry out such detailed specification and presentation of statistical distributions. Yet, the power of computing hardware has increased enormously in the last few years. For example, many large statistical analysis programs now incorporate subroutines to provide p-values for the F-ratio or for the chi-square statistic. These routines are provided within the programs as a convenience, to save the user the effort of looking up the values in a table. This kind of idea, now used only incidentally, can be extended considerably. Particularly Bayesian data analysis, but also conventional analysis, can thus be made more efficient if a computer routine is used to generate representational results tailored to a specific problem, rather than relying on published tables which are incomplete and partly inappropriate. Such computer routines now are quite feasible for machines which cost a few thousand dollars or less, and fit on a desktop.

In conjunction with the clear feasibility of such routines, the limitations of existing packages (reviewed just previously) indicate that some additional programs aimed to close some of these gaps will be valuable. Specifically, it would be desirable to have a computer program capable of providing detailed description of _any_ Student's _t_ distribution (not just the standardized one), and for any reasonable number of degrees of freedom. Such a program should also provide corresponding results for the Behrens-Fisher distribution, since this is encountered in many situations parallel to those involving the Student's _t_. However, a program for the Behrens-Fisher distribution will be especially useful because published tables of that distribution are scarce, with the existing ones providing only coarse coverage (Pearson and Hartley, 1966; Fisher and Yates, 1957).

A computer program of this sort, ideally, should provide the density value and the cumulative probability at each of a number of points specified by the user. Also, the program should be interactive and flexible, enabling the user to control pertinent aspects of his analysis conveniently (i.e., graphic and/or tabular output, labeling, ranges of values considered, etc.). Finally, the program should be understandable to non-expert programmers and should be available in a language which is widely implemented.

3.4 Programming the Student's t and Behrens-Fisher Distributions

Development of a computer program with all of these features has been relatively difficult, primarily because the Behrens-Fisher distribution, unlike Student's t distribution, is computationally complex. In general, the direct calculation of a single value along the density curve of the Behrens-Fisher distribution requires the numerical evaluation of a messy integral. Thus, evaluation of the Behrens-Fisher distribution is a time-consuming procedure, even with modern computers, and so is impractical for a program which is to be run routinely. (But see Fisher and Healy, 1956; Patil, 1963 for routines which for certain cases do provide highly accurate results on the Behrens-Fisher distribution, using analytic formulas.)

To circumvent this computational obstacle for the general case, several writers have investigated various approximations to the

Behrens-Fisher distribution. Perhaps the best-known results (Brownlee, 1965; Snedecor and Cochran, 1967; Winer, 1971) are due to Welch (1938, 1947, 1949) and to Aspin (1949). These have been tabulated in Pearson and Hartley (1966, Table 11). However, a less well-known but excellent approximation has been developed by V.K. Patil (Patil, 1963, 1964a, 1964b, 1965), working from premises consistent with the approach of Behrens, Fisher, and Jeffreys. Patil's approximation used a "dilated" Student's t distribution to approximate the Behrens-Fisher distribution. The logic of the approximation is to match the second and fourth moments of the two distributions (the Student's t and the Behrens-Fisher); this implies that the first four moments are matched, since the odd moments of both distributions are zero (see also Tanburn, 1938). Patil's approximation appears to be quite accurate, and to be entirely satisfactory for use in most situations, although the documentation presently available on it is not as complete as one would wish.

To be more specific, let us use f_1 and f_2 to denote the degrees of freedom in each of the two groups under consideration, and d to denote the value of the standardized Behrens-Fisher variable. That is, d is the variable which follows a Behrens-Fisher distribution with mean equal to zero and scale factor equal to unity. Patil states that her approximation has a maximum error of less than 7.42% for any density value, provided that both f_1 and f_2 are greater than or equal to 7, and that the absolute value of d is less than 5.0 (Patil, 1965: 269). Most of the approximated values, particularly those near the center of the Behrens-Fisher distribution, are considerably more accurate. The accuracy also increases quickly as the number of degrees of freedom increases. Typically, the density values provided by the approximation differ by less than 1.0% from the true density values (Patil, 1965). An additional and very convenient property of this approximation algorithm is that it can be quite easily included as part of a computer program which calculates the densitites and cumulative probabilities for the Student's t distribution.

Operationally, then, a program can be designed so that it always calculates a Student's t density and cumulative probability. In some situations, those results will be used as the approximating results for a Behrens-Fisher distribution. There is, however, one technical complication in using this approach to the evaluation of the Behrens-Fisher distribution. To carry out the approximation, one must calculate the parameters of the Student's t distribution (i.e., the mean, the degrees of freedom, and the scale factor) according to formulas developed by Patil (Patil, 1965; Box and Tiao, 1973: 107). Those formulas are reasonably simple, but they often lead to non-integer degrees of freedom for the approximating t distribution. This is similar to the situation with the more widely used Welch-Aspin approximation formula. Therefore, any algorithm for the Student's t which is also to be used as an approximating distribution for the Behrens-Fisher must properly handle such cases. This requirement creates no problem for the density calculation, but does require a special form for the calculation of the cumulative probabilities.

The concrete task remaining, then, is to choose an acceptable algorithm for the cumulative probability of Student's t. Several fairly well-known algorithms exist. For instance, a series expansion of the standardized Student's t distribution is available (Abramowitz and Stegun, 1968:948, eqs 26.7.3 and 26.7.4; Johnson and Kotz, vol. 2, 1970:96, eqs 4.3 and 4.4) which provides theoretically exact values of the cumulative probabilities. Although this algorithm is exact, it has two drawbacks for our purposes. First and most important, it is applicable only if the degrees of freedom for the t distribution is an integer. Second, the computing effort involved in this formula is proportional to the degrees of freedom. This means that for large degrees of freedom the formula will be quite slow and also subject to the buildup of rounding errors in any finite-length calculation device. These drawbacks, especially that of the time required, were explored using a 13-digit computer (Wang 2200), and were found to become serious if the degrees of freedom are greater than approximately 15 to 30. This is especially true of the decrease in speed.

A second algorithm, described in Johnson and Kotz (1970, vol. 2: 101-102, eq 10) is derived from work by R. A. Fisher (1925). This formula is relatively simple and is capable of handling non-integer degrees of freedom. Also, its computing time is independent of the degrees of freedom or other parameters of the t distribution. However, as indicated by Fisher (1925), the accuracy of this formula is inadequate when the degrees of freedom are small. A comparison was carried out of this formula as implemented on the Wang 2200 with the cumulative probabilities of Student's t given in Table 9 of Pearson and Hartley (1966). That comparison indicates that the accuracy of this Fisher formula is indeed poor for degrees of freedom below about 9 or 10, and is only fair until about 18 degrees of freedom. Above 18 degrees of freedom, the Fisher formula agrees perfectly, as Fisher noted, with the published table entries.

Several published algorithms for the Student's t distribution (cf. Morris, 1968; Levine, 1969; Hill, 1970) have been constructed as hybrids of the series expansion mentioned above for small degrees of freedom, and of the Fisher formula (or similar formulas) for larger degrees of freedom (e.g., above 20). But, notice that none of these algorithms can provide answers for the cumulative probabilities when the degrees of freedom is small and is also a non-integer.

There is, however, another approach available. The Student's t distribution can be expressed as a special case of the incomplete Beta function ratio (Johnson and Kotz, vol. 2, 1970; Abramowitz and Stegun, 1968; Welch, 1938; Soper, 1921). Once this is recognized, the problem of evaluating a cumulative Student's t becomes a special case of the problem of evaluating an incomplete Beta function ratio. This latter approach has been illustrated by Majumder and Bhattacharjee (1973). Prior to their paper, however, the approach was adopted, in an elegant and efficient way, in the MANECON package. That is, the MANECON procedure for evaluating a cumulative Student's t makes use of the incomplete Beta function ratio algorithm. A simplified version of the MANECON algorithm (removing aspects of the algorithm not needed as long as it is used only for evaluating a

Student's t) was programmed for the Wang 2200 and then compared in detail with the contents of Table 9 in Pearson and Hartley (1966). The agreement found was excellent throughout. Specifically, Pearson and Hartley's Table 9 gives cumulative probabilities for Student's t for degrees of freedom 1(1)23, 30, 40, 60, 120, and infinity. The t values are reported from 0(.1)4, plus additional values for some degrees of freedom. In short, this table is the most complete and authoritative source available for the cumulative Student's t. For both the Wang 2200 version written in BASIC, and for a DECsystem-10 version written in FORTRAN, the agreement between the MANECON-modified algorithm and the Pearson and Hartley table was within ± .0001, and usually within ± .00005. This excellent agreement could have been further improved if a convergence boundary parameter in the MANECON algorithm had been adjusted for greater accuracy, but this would have led to greater times for the computation. The accuracy found here was uniform for all degrees of freedom, including very small and very large. Finally, the computing time required by this algorithm is reasonably short, corresponding to the time for the series expansion in the range of 20 to 25 degrees of freedom. Therefore, on all criteria, this modified MANECON algorithm is the best for our needs, and meets them fully.

Once the accuracy of the algorithm had been established, a new program was developed around it to evaluate the Student's t and the Behrens-Fisher distributions. Permission was obtained from the President and Fellows of Harvard College for this adaptation of the MANECON algorithm. The logic of the algorithm is described in detail in the book Computer Programs for Elementary Decision Analysis, by Robert Schlaifer (1971:200-205).

This new program has been named "TBF". It provides, for either a Student's *t* or a Behrens-Fisher distribution, a list of density values and cumulative probabilities, as well as a graph of the density curve. The program has been implemented in three dialects of the BASIC language, and one version of FORTRAN. It is currently operational on the Wang 2200, the GE-635 using the Dartmouth Time Sharing Services BASIC, and the DECsystem-10 in BASIC. It is also operational on the DECsystem-10 in FORTRAN. The coding of the BASIC versions is reasonably standard with the exception of certain machine-specific input-output declarations, and can be easily adapted to almost any machine for which a BASIC language exists. The FORTRAN version likewise uses only standard instructions in the calculations, and a few machine-specific instructions for control of input-output devices, etc. All the versions are nearly identical in operation. The BASIC versions allow use of the program on micro-processor machines (such as the Wang 2200, the Hewlett-Packard 9830, the IBM 5100, or the Altair 8800) which do not have a FORTRAN compiler, and the FORTRAN version will allow more economical execution on larger computers. A listing of the Wang 2200 version of TBF is given in Figure 3.2.

As previously noted, the algorithm from the MANECON package used in TBF provides excellent accuracy for the cumulative Student's *t* distribution. There is, however, only sketchy documentation available in the statistical literature for the accuracy of Patil's approximation of the cumulative Behrens-Fisher distribution. That

Figure 3.2

```
76/11/15                     TBF                     PAGE 1
PROGRAM LISTING

 1010  REM "TBF" DENSITY & CUMUL FOR -T- OF BEHRENS DISTRIBUTION
 1020  REM JAMES FENNESSEY & SUSAN RADIUS
 1030  REM (SOCIAL RELATIONS DEPT., JOHNS HOPKINS UNIVERSITY)
 1040  REM VERSION 3: REVISED 15 NOV 76  2200
 1050  DIM A$64
      :SELECT PRINT 005(64)
      :PRINT HEX( 03)
 1060  T9=.00005
      :T8=200
      :T7=227
      :T6=125
      :D9=4000
      :R4=4
 1070  F1=1.330274429
      :F2=1.821255978
      :F3=1.781477937
 1080  F4=0.356563782
      :F5=0.319381530
      :F6=SQR( 2*#PI )
 1090  INPUT "TYPE 2 FOR PLOT ONLY; 1 FOR OUTPUT ON PAPER",S9
 1100  IF S9<>1  THEN 1110
      :SELECT PRINT 015(80)
      :PRINT
 1110  PRINT "'TBF' DISTRIBUTION EVALUATOR"
      :GOSUB 2030
 1120  SELECT PRINT 005(64)
 1130  PRINT "TYPE LABEL FOR PROBLEM (TO STOP, TYPE 'END')"
 1140  INPUT A$
      :IF S9<>1  THEN 1150
      :SELECT PRINT 015(80)
 1150  IF A$="END"  THEN 1670
      :PRINT A$
 1160  INPUT "TYPE 1 IF SINGLE T, 2 IF DUAL T, 3 IF BEHRENS",S8
 1170  ON S8 GOTO 1180  , 1200 , 1200
 1180  INPUT "TYPE N, MEAN, VARIANCE",N,M,V
      :D=N-1
      :S2=V/N
      :REM D=D.F.;S2=SCALE↑2
 1190  L$="SINGLE -T-"
      :PRINT L$," N"," M"," V"
      :PRINT ,N,M,V
      :GOTO 1320
 1200  INPUT "TYPE N, MEAN, VARIANCE OF GROUP 1",N1,M1,V1
 1210  INPUT "TYPE N, MEAN, VARIANCE OF GROUP 2",N2,M2,V2
 1220  D1=N1-1
      :D2=N2-1
      :L$="BEHRENS"
      :IF S8=3  THEN 1260
      :L$="DUAL -T-"
 1230  D=D1+D2
      :M=M2-M1
```

Figure 3.2(continued)

```
76/11/15                        TBF                        PAGE 2
PROGRAM LISTING

      :S2=(1/N1+1/N2)*(D1*V1+D2*V2)/D
      :GOTO 1310
 1240  REM SEE G.E.P. BOX & G.C. TIAO, BAYESIAN INFERENCE IN
 1250  REM STATISTICAL ANALYSIS (ADDISON-WESLEY, 1973) P. 107
 1260  C9=(V2/N2)/(V1/N1+V2/N2)
      :C8=1-C9
      :C7=D2/(D2-2)*C9+D1/(D1-2)*C8
 1270  C6=D2↑2/(D2-2)↑2/(D2-4)*C9↑2
      :C5=D1↑2/(D1-2)↑2/(D1-4)*C8↑2
 1280  C4=C6+C5
      :C3=V1/N1+V2/N2
      :D=4+C7↑2/C4
      :M=M2-M1
      :S2=(D-2)/D*C7*C3
 1290  S7=SQR( V1/N1*N2/V2)
      :PRINT "PHI=";180/#PI *ARC TAN( S7);"  TAN(PHI)=";S7
 1300  REM END OF BEHRENS PREPARATION
 1310  PRINT L$," N"," M"," V"
      :PRINT "GRP1",N1,M1,V1
      :PRINT "GRP2",N2,M2,V2
 1320  S=SQR( S2)
      :REM S= -T- SCALE FACTOR
 1330  PRINT
      :IF S8<>3  THEN 1340
      :PRINT " FOR -T- DIST USED AS APPROXIMATION,"
 1340  PRINTUSING 1350  ,FN R(D),FN R(M),FN R(S)
 1350  % D.F. = #####.####;   MEAN = -###.####;   SCALE FACTOR = ###.####
 1360  IF D>D9  THEN 1390
 1370  REM CALCULATE K
      :G9=(D+1)/2
      :GOSUB 1700
      :S5=G0
      :G9=D/2
      :GOSUB 1700
      :S4=G0
 1380  K=S5-(S4+LOG( S*SQR( #PI *D)))
      :REM K NOW READY
 1390  SELECT PRINT 005(64)
      :L=M-(3.5*S)
      :U=L+2*(M-L)
 1400  X=L
      :P=FN T(X)
      :IF ABS( P)<T7  THEN 1410
      :P=SGN( P)*T7
 1410  F=EXP( P)
 1420  IF D<D9  THEN 1430
      :GOSUB 2030
      :GOTO 1440
 1430  GOSUB 2090
      :REM F=DENSITY, C=CUM PROB
 1440  PRINT "L = ";FN R(L);"U = ";FN R(U)
```

Figure 3.2(continued)

```
76/11/15                         TBF                             PAGE 3
PROGRAM LISTING

      :PRINTUSING 1450  ,FN R(L),FN R(P),FN R(C)
 1450  %  AT L = -###.####, DENSITY = #.####  CUMUL PROB = #.####
 1460  PRINT "TYPE NEW LOWER BOUND OR ELSE TYPE ";M
      :INPUT L1
 1470  IF ABS( L1-M)<.00005  THEN 1500
      :IF L1>M  THEN 1490
 1480  INPUT "TYPE NEW UPPER BOUND",U
      :L=L1
      :GOTO 1400
 1490  PRINT "L MUST BE < M"
      :GOTO 1460
 1500  IF S9<>1  THEN 1510
      :SELECT PRINT 015(80)
 1510  PRINT "L = ";L,"  U = ";U
 1520  INPUT "WIDTH OF AN INTERVAL",W
      :W1=INT( (U-L)/W+.00001)+1
 1530  IF W1<=T8  THEN 1540
      :PRINT "TOO MANY POINTS"
      :GOTO 1520
 1540  PRINT "VALUES FOR ";W1;" POINTS."
 1550  IF S9<>2  THEN 1560
      :GOSUB 1820
      :REM PLOTS DENSITY
      :GOTO 1050
 1560  PRINT
      :PRINT "        X-AXIS","    ";"DENSITY","CUMUL. PROB."
      :PRINT
 1570  FOR J=1  TO W1
      :X=L+(J-1)*W
      :P=FN T(X)
 1580  IF ABS( P)<T7  THEN 1590
      :P=SGN( P)*T7
 1590  P=EXP( P)
 1600  IF D<D9  THEN 1610
      :GOSUB 2330
      :GOTO 1620
 1610  GOSUB 2090
 1620  PRINTUSING 1630  ,J,FN R(X),FN R(P),FN R(C)
      :NEXT J
 1630  % ### -###.####     ###.####       ##.####
 1640  PRINT
      :PRINT "END OF TABULATION"
      :GOSUB 2030
 1650  INPUT "DO YOU WANT THE PLOT",B$
      :IF STR( B$,1,1)="Y"  THEN 1660
      :GOTO 1050
 1660  GOSUB 1800
      :REM PLOTS DENSITY
      :PRINT
      :GOTO 1050
 1670  STOP
```

Figure 3.2(continued)

```
76/11/15                    TBF                         PAGE 4
PROGRAM LISTING

     :REM END OF MAIN PROGRAM
1680    REM ***************************
1690    REM SETS G0=LN(GAMMA(G9)), USING STIRLING FORMULA
1700  IF G9<=1E30  THEN 1710
     :G0=1E38
     :RETURN
1710  IF G9>1E-9  THEN 1720
     :G0=0
     :RETURN
1720  IF G9<1E10  THEN 1730
     :G0=G9*(LOG( G9)-1)
     :RETURN
1730  G8=G9
     :G7=1
1740  IF G8>18  THEN 1750
     :G7=G7*G8
     :G8=G8+1
     :GOTO 1740
1750  G6=1/G8↑2
     :G5=1/12
     :G4=1/360
     :G3=1/1260
     :G2=1/1680
     :G1=LOG( SQR( 2*#PI ))
1760  G0=G1+(G8-.5)*LOG( G8)-G8+1/G8*(G5-(G6*(G4+(G6*(G3-G6*G2)))))-LOG( G7)
1770  RETURN
1780    REM
1790    REM PLOTS Y VERSUS X
1800  INPUT "TYPE LOWER AND UPPER BOUND AND STEP WIDTH",L,U,W
     :W1=INT( (U-L)/W+1)
1810  IF W1<=T8  THEN 1820
     :PRINT "TOO MANY POINTS"
     :GOTO 1800
1820  P=FN T(M)
     :IF ABS( P)<T7  THEN 1830
     :P=SGN( P)*T7
1830  P=EXP( P)
     :Y9=P/50
     :REM Y9=ONE PRINT POS
1840  SELECT PRINT 015(80)
     :PRINT
     :GOSUB 2030
     :PRINT A$
1850  PRINTUSING 1860  ,L$,FN R(D),FN R(M),FN R(S)
1860  % ########## DENSITY ON #####.#### D.F., MEAN = -###.####, SCALE  = ###
     #.####
1870  Y8=INT( Y9*10↑4+.5)/10↑4
1880  PRINTUSING 1890  ,Y8
1890  % X-AXIS IS * IF DENSITY < ##.####; OTHERWISE IS I
1900  PRINTUSING 1910  ,10*Y8
1910  %                     TICK MARKS ON Y-AXIS EVERY ###.####
```

```
                                                    PAGE 5
76/11/15                   TBF
PROGRAM LISTING

 1920  PRINT
      :PRINT TAB( 16);"0";TAB( 64);50*Y8
 1930  PRINT TAB( 16);"I----:----I----:----I----:----I----:----I----:----I----
      :"
 1940  FOR J=1  TO W1
      :B$="*"
      :C$="I"
      :X=L+(J-1)*W
      :F=FN T(X)
 1950  IF ABS( F)<T7  THEN 1960
      :F=SGN( F)*T7
 1960  P=EXP( F)
      :IF P>=Y8  THEN 1970
      :B$=" "
      :C$="*"
 1970  Y=INT( P/Y9+.5)
      :X=INT( X*10↑4+.5)/10↑4
 1980  PRINT J;TAB( 6);
      :PRINTUSING 2000  ,FN R(X);
 1990  PRINT TAB( 16);C$;TAB( 16+Y);B$
      :NEXT J
 2000  % -###.####
 2010  PRINT
      :GOSUB 2030
      :PRINT
      :SELECT PRINT 005(64)
      :RETURN
 2020    REM
 2030  PRINT "--------------------------------------------------"
      :RETURN
 2040    REM
 2050  DEFFN T(X)=K+LOG( (1+(((X-M)↑2)*(D+1))/(D*S2*(D+1)))↑(-(D+1)/2))
 2060    REM GETS CUMUL -T- PROBABILITY
 2070  REM (FROM HARVARD BUSINESS SCHOOL 'MANECON' PACKAGE, WITH PERMISSION)
 2080  REM INPUTS X,M,S,D; OUTPUT C
 2090  T5=.5*D
      :T=(X-M)/S
      :Z2=.5*(1+T/SQR( D+T*T))
 2100  Z=Z2
      :IF Z2<.5  THEN 2110
      :Z=1-Z2
 2110  IF Z<0  THEN 2300
      :IF Z>1  THEN 2300
      :IF T5<0  THEN 2300
      :IF A0=T5  THEN 2130
 2120  G9=T5
      :GOSUB 1700
      :D3=G0
      :G9=D
      :GOSUB 1700
      :D4=G0
```

Figure 3.2(continued)

```
76/11/15                          TBF                          PAGE 6
PROGRAM LISTING

      :A0=T5
 2130  A5=(T5-1)*LOG( (1-Z)*Z)-2*D3+D4
 2140  IF A5<-T7  THEN 2300
      :IF A5<=T7  THEN 2150
      :A5=T7
 2150  A2=EXP( A5)
 2160  REM INITIALIZE FOR CUMUL
      :C=0
      :A6=Z/(1-Z)
      :A7=A2*(1-Z)/T5
 2170  REM EVALUATE BINOMIAL SERIES
 2180  E1=INT( T5)
      :IF E1<1  THEN 2230
      :E2=T6
      :IF E1>T6  THEN 2190
      :E2=E1
 2190  A8=0
      :FOR I=1  TO E2
      :E3=(T5-A8)/(T5+A8)*A6
 2200  IF A7>T9  THEN 2210
      :IF A7<(T9/E3-T9)  THEN 2280
 2210  A7=E3*A7
      :C=C+A7
      :A8=A8+1
      :NEXT I
      :IF E2<E1  THEN 2270
 2220  REM EVALUATE POWER SERIES
 2230  E4=E1
      :IF E4=T5  THEN 2280
      :A7=A7*(T5-E4)/(T5+E4)*Z/(1-Z)↑(T5-E4)
 2240  C=C+A7
      :E5=T9/A6
      :A8=E4+1
      :E6=E1+1
      :FOR I=E6  TO T6
 2250  A7=(T5+A8-1)/(T5+A8)*(A8-T5)/(A8-E4)*Z*A7
 2260  C=C+A7
      :IF A7<E5  THEN 2280
      :A8=A8+1
      :NEXT I
 2270  PRINT "FAILED TO CONVERGE: CUM PROB MAY BE INACCURATE"
 2280  IF Z2<.5  THEN 2290
      :C=1-C
 2290  RETURN
      :REM NORMAL RETURN: C = CUMUL PROBAB
 2300  C=1
      :IF Z2<.5  THEN 2310
      :C=1-C
 2310  RETURN
 2320    REM EVALUATES CUMULATIVE NORMAL DISTRIB
 2330  Y3=(X-M)/S
```

Figure 3.2(continued)

```
                                TBF                                   PAGE 7
76/11/15
PROGRAM LISTING

        :Y4=ABS( Y3)
 2340   F7=1/F6*EXP( -.5*Y3↑2)
        :F8=1/(1+.2316419*Y4)
 2350   C=1-F7*F8*((((F1*F8-F2)*F8+F3)*F8-F4)*F8+F5)
 2360   IF X>M  THEN 2370
        :C=1-C
 2370   RETURN
 2380   DEFFN R(X)=SGN( X)*INT( ABS( X)*10↑R4+.5)/10↑R4
```

is, Patil carried out a check of the approximated *density* values for *odd* degrees of freedom in some detail, but provided only a few general remarks concerning the limits of accuracy for the cumulative approximation. Obviously, if the density values are highly accurate, then the cumulative distribution must also be reasonably accurate. But, when the degrees of freedom are small, and the accuracy of the densities begins to degrade, there is no knowledge as to the behavior of the cumulative approximation. The situation is most serious for small, *even* degrees of freedom, since no analytical formula for the Behrens-Fisher distribution is available for those cases. The problem appears likely to be most serious also for values of *d* in the tails of the Behrens-Fisher distribution, or when the angle "phi" used in that distribution (see Box and Tiao, 1973) is near 45 degrees.

3.5 Exploring the Utility of Numerical Integration

To clarify the limits of the approximation's utility, and at the same time to explore the behavior of alternative means of numerical integration of a symmetric curve such as that characterizing Student's *t*, some direct "brute force" numerical integration solutions of the Behrens-Fisher cumulative distribution were developed and programmed on the Wang 2200. The integration was approached via several widely used methods, and for both small and large degrees of freedom. The logic behind this is that the accuracy of the results obtained when the degrees of freedom were large would provide clues to the quality of the numerical integration formulas, and the range of panel sizes that might be needed; the formulas found to be best then could be used to develop a direct calculation of the Behrens-Fisher cumulative solution when the degrees of freedom are small, and when the angle is near 45 degrees. This work has not yet been fully completed in all its interesting ramifications, but some major conclusions are clear. The accuracy of Patil's approximation for the cumulative Behrens-Fisher distribution is within about 5% for cases in which both degrees of freedom of the Behrens-Fisher distribution are greater than or equal to 7, and in which the standardized Behrens-Fisher variable _d_ is less than 5 in absolute value. If the degrees of freedom decrease, the accuracy of the approximation declines smoothly and is within about 6% if both degrees of freedom are greater than or equal to 6, and if the variable _d_ is less than or equal to 3 in absolute value. For extremely small samples, the accuracy degrades more quickly, and most seriously in the tails, but no simple summary is possible. However, our verification check clearly indicates that Patil's approximation is suitable for research uses as long as the degrees of freedom for the two groups are not _very_ small.

The other focus of this investigation was to ascertain the advantages (if any) of different approaches to numerical integration, in terms of accuracy per computing time, and stability across different situations. Our results in this respect were derived by programming a general purpose numerical integration illustrator program, and running it with several functions. This program, which we have labeled "INTEG", incorporates five separate algorithms of general interest for the calculation of numerical integrals: (1)

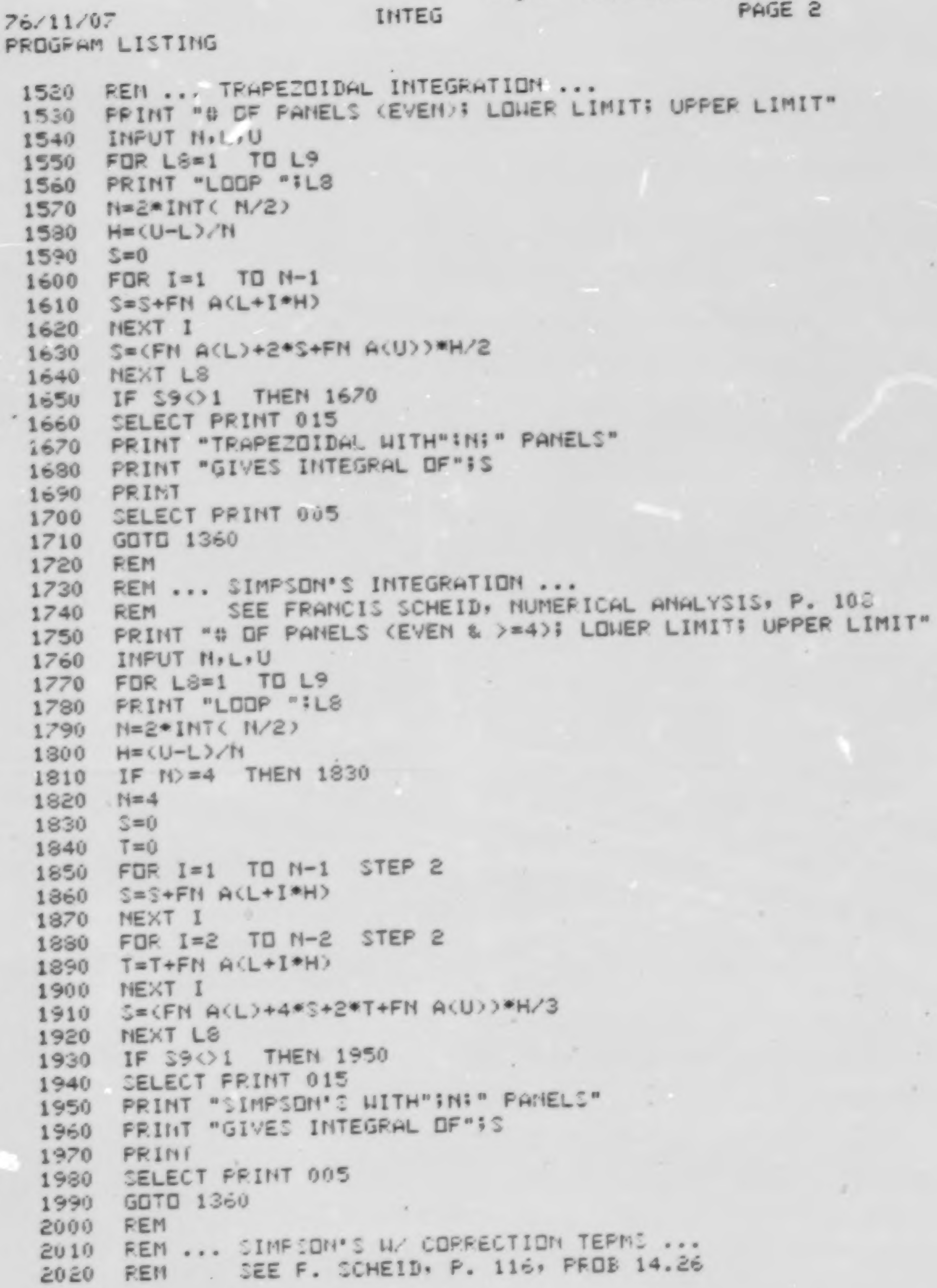
composite trapezoidal rule; (2) composite Simpson' rule; (3) composite Simpson's rule with correction terms; (4) Romberg integration; and (5) Gauss-Legendre quadrature using a variable number of points. A listing of this program in one version (employing 16 point Gaussian quadrature) is given as Figure 3.3. The results from executing this program indicated that the Gaussian and the Simpson's rule with correction terms were the most accurate methods for a given amount of computing time, and thus are generally preferable to the other methods for most uses. The program has been developed so that the function being evaluated is represented as FNA(X), and can be replaced rather easily. In the sample listing shown in Figure 3.3, the standardized Normal density function has been used as a simple illustration. The results for this function and for others indicate that the usual notion that the Gaussian quadrature is optimal may be modified somewhat if the Simpson's with correction terms is considered as an alternative.

A second program, known as "BFTEST", also was developed using the Simpson's rule formulas, the Gaussian quadrature, and others. This program was used to carry out the exact numerical checking of the accuracy of Patil's approximation, as well as to compare alternative integration procedures. One peculiarity emerging from the trial runs made with the Gaussian quadrature was that, on the particular function involved in the Behrens-Fisher formulas, the Gaussian integration was not quite as accurate as the Simpson's rule. This appeared to depend on the range of values over which the inner integration was done. This finding takes us quite far from our

Figure 3.3

```
76/11/07                      INTEG                      PAGE 1
PROGRAM LISTING

 1010  REM "INTEG"  NUMERICAL INTEGRATION DEMONSTRATOR ROUTINE
 1020  REM PROGRAMMED BY JAY ROCKSTROH
 1030  REM 07 NOV 76  1420
 1040  REM DEFINE FUNCTION TO BE INTEGRATED AS FNA(X)
 1050  REM
 1060  REM FOLLOWING IS STANDARDIZED NORMAL DENSITY FORMULA
 1070  DEFFN A(X)=1/SQR( 2*3.14159265359)*EXP( -.5*X*X)
 1080  REM
 1090  REM L9=# OF LOOPS (USED FOR GETTING TIMINGS)
 1100  REM INTEGRATION TYPES AVAILABLE ARE -
 1110  REM 1=TRAPEZOID
 1120  REM 2=SIMPSON'S
 1130  REM 3=SIMPSON'S W/ CORRECTION TERMS
 1140  REM 4=ROMBERG
 1150  REM 5=GAUSS-LEGENDRE QUADRATURE (16 POINT)
 1160  REM TYPE 6 TO END THIS JOB
 1170  REM
 1180  REM ...INITIALIZATION...
 1190  DIM X(8),W(8),A$(5)30
 1200  DIM T(20,20)
 1210  MAT READ A$
 1220  DATA "TRAPEZOIDAL","SIMPSON'S"
 1230  DATA "SIMPSON'S W/ CORRECTIONS","ROMBERG","GAUSS-LEGENDRE"
 1240  FOR I=1  TO 8
 1250  READ X(I),W(I)
 1260  NEXT I
 1270  DATA .9894009350, .0271524594
 1280  DATA .9445750231, .0622535239
 1290  DATA .8656312024, .0951585117
 1300  DATA .7554044084, .1246289713
 1310  DATA .6178762444, .1495959888
 1320  DATA .4580167777, .1691565194
 1330  DATA .2816035508, .1826034150
 1340  DATA .0950125098, .1894506105
 1350  REM
 1360  SELECT PRINT 005(64)
 1370  INPUT "TYPE 1 FOR OUTPUT ON PRINTER",S9
 1380  PRINT "# OF METHOD"
 1390  INPUT I
 1400  IF I<>6  THEN 1450
 1410  PRINT
 1420  PRINT
 1430  PRINT "END OF JOB"
 1440  STOP
 1450  L9=1
 1460  REM DEFAULT VALUE OF L9 (# OF LOOPS)
 1470  INPUT "# OF LOOPS (DEFAULT = 1)",L9
 1480  PRINT A$(I)
 1490  REM
 1500  ON I GOTO 1520  , 1730 , 2010 , 2410 , 2830 , 1410
 1510  REM
```

Figure 3.3(continued)



```
76/11/07                     INTEG                      PAGE 2
PROGRAM LISTING

 1520  REM ... TRAPEZOIDAL INTEGRATION ...
 1530  PRINT "# OF PANELS (EVEN); LOWER LIMIT; UPPER LIMIT"
 1540  INPUT N,L,U
 1550  FOR L8=1  TO L9
 1560  PRINT "LOOP ";L8
 1570  N=2*INT( N/2)
 1580  H=(U-L)/N
 1590  S=0
 1600  FOR I=1  TO N-1
 1610  S=S+FN A(L+I*H)
 1620  NEXT I
 1630  S=(FN A(L)+2*S+FN A(U))*H/2
 1640  NEXT L8
 1650  IF S9<>1  THEN 1670
 1660  SELECT PRINT 015
 1670  PRINT "TRAPEZOIDAL WITH";N;" PANELS"
 1680  PRINT "GIVES INTEGRAL OF";S
 1690  PRINT
 1700  SELECT PRINT 005
 1710  GOTO 1360
 1720  REM
 1730  REM ... SIMPSON'S INTEGRATION ...
 1740  REM      SEE FRANCIS SCHEID, NUMERICAL ANALYSIS, P. 108
 1750  PRINT "# OF PANELS (EVEN & >=4); LOWER LIMIT; UPPER LIMIT"
 1760  INPUT N,L,U
 1770  FOR L8=1  TO L9
 1780  PRINT "LOOP ";L8
 1790  N=2*INT( N/2)
 1800  H=(U-L)/N
 1810  IF N>=4  THEN 1830
 1820  N=4
 1830  S=0
 1840  T=0
 1850  FOR I=1  TO N-1  STEP 2
 1860  S=S+FN A(L+I*H)
 1870  NEXT I
 1880  FOR I=2  TO N-2  STEP 2
 1890  T=T+FN A(L+I*H)
 1900  NEXT I
 1910  S=(FN A(L)+4*S+2*T+FN A(U))*H/3
 1920  NEXT L8
 1930  IF S9<>1  THEN 1950
 1940  SELECT PRINT 015
 1950  PRINT "SIMPSON'S WITH";N;" PANELS"
 1960  PRINT "GIVES INTEGRAL OF";S
 1970  PRINT
 1980  SELECT PRINT 005
 1990  GOTO 1360
 2000  REM
 2010  REM ... SIMPSON'S W/ CORRECTION TERMS ...
 2020  REM      SEE F. SCHEID, P. 116, PROB 14.26
```

Figure 3.3(continued)

```
                                                        PAGE 3
76/11/07                       INTEG
PROGRAM LISTING

 2030  PRINT "# OF PANELS (EVEN & >=8); LOWER LIMIT; UPPER LIMIT"
 2040  INPUT N,L,U
 2050  FOR L8=1  TO L9
 2060  PRINT "LOOP ";L8
 2070  N=2*INT( N/2)
 2080  H=(U-L)/N
 2090  IF N>=8  THEN 2110
 2100  N=8
 2110  M1=0
 2120  M2=0
 2130  FOR I=3  TO N-3  STEP 2
 2140  M1=M1+FN A(L+I*H)
 2150  NEXT I
 2160  FOR I=4  TO N-4  STEP 2
 2170  M2=M2+FN A(L+I*H)
 2180  NEXT I
 2190  T1=FN A(L-H)
 2200  T2=FN A(L)
 2210  T3=FN A(L+H)
 2220  T4=FN A(L+2*H)
 2230  T5=FN A(U-2*H)
 2240  T6=FN A(U-H)
 2250  T7=FN A(U)
 2260  T8=FN A(U+H)
 2270  S0=T2+4*T3+2*T4+4*M1+2*M2+2*T5+4*T6+T7
 2280  S4=T1-4*T2+7*T3-8*T4+8*M1-8*M2-8*T5+7*T6-4*T7+T8
 2290  S6=FN A(L-2*H)-6*T1+16*T2-26*T3+31*T4-32*M1
 2300  S6=S6+32*M2+31*T5-26*T6+16*T7-6*T8+FN A(U+2*H)
 2310  S=S0*H/3-S4*H/90+S6*H/756
 2320  NEXT L8
 2330  IF S9<>1  THEN 2350
 2340  SELECT PRINT 015
 2350  PRINT "SIMPSON'S W/ CORRECTION WITH";N;" PANELS"
 2360  PRINT "GIVES INTEGRAL OF";S
 2370  PRINT
 2380  SELECT PRINT 005
 2390  GOTO 1360
 2400  REM
 2410  REM ... ROMBERG INTEGRATION ...
 2420  REM      SEE CARNAHAN, LUTHER, & WILKES, P 90-99, ESPEC. 96
 2430  PRINT "MAX # OF PANELS (EVEN); LOWER LIMIT; UPPER LIMIT"
 2440  INPUT N,L,U
 2450  N2=INT( LOG( N)/LOG( 2)+.00005)
 2460  PRINT "NMAX=";N2
 2470  E9=0
 2480  INPUT "ERROR TOLERANCE (DEFAULT = 0)",E9
 2490  FOR L8=1  TO L9
 2500  PRINT "LOOP ";L8
 2510  MAT T=ZER(N2+1,N2+1)
 2520  H1=U-L
 2530  I9=0
 2540  T(1,1)=(FN A(L)+FN A(U))*H1/2
```

Figure 3.3(continued)

```
76/11/07                  INTEG                 PAGE 4
PROGRAM LISTING

 2550   REM HALVE INTERVAL, COMPUTE T(N1+1,1)
 2560   FOR N1=1  TO N2
 2570   T(N1+1,1)=0
 2580   H1=H1/2
 2590   I9=I9*2+1
 2600   FOR I=1  TO I9  STEP 2
 2610   T(N1+1,1)=T(N1+1,1)+FN A(L+I*H1)
 2620   NEXT I
 2630   T(N1+1,1)=T(N1,1)/2+T(N1+1,1)*H1
 2640   REM FILL IN TABLEAU
 2650   F2=1
 2660   FOR J=2  TO N1+1
 2670   F2=4*F2
 2680   K=N1+2-J
 2690   T(K,J)=(F2*T(K+1,J-1)-T(K,J-1))/(F2-1)
 2700   NEXT J
 2710   T9=ABS( T(1,N1)-T(1,N1+1))
 2720   IF T9<E9  THEN 2740
 2730   NEXT N1
 2740   NEXT L8
 2750   IF S9<>1  THEN 2770
 2760   SELECT PRINT 015
 2770   PRINT "ROMBERG  ( N2=";N2;"  N1=";N1;
 2780   PRINT "   ERROR TOLERANCE=";E9;")"
 2790   PRINT "GIVES INTEGRAL OF";T(1,N1+1)
 2800   PRINT
 2810   SELECT PRINT 005
 2820   GOTO 1360
 2830   REM
 2840   REM ... GAUSS-LEGENDRE QUADRATURE ...
 2850   PRINT "LOWER LIMIT; UPPER LIMIT"
 2860   INPUT L,U
 2870   FOR L8=1  TO L9
 2880   PRINT "LOOP ";L8
 2890   S=0
 2900   FOR I=1  TO 8
 2910   X1=( (U-L)*X(I)+U+L )/2
 2920   X2=(-(U-L)*X(I)+U+L )/2
 2930   S=S+W(I)*(FN A(X1)+FN A(X2))
 2940   NEXT I
 2950   S=S*(U-L)/2
 2960   NEXT L8
 2970   IF S9<>1  THEN 2990
 2980   SELECT PRINT 015
 2990   PRINT "GAUSS-LEGENDRE QUADRATURE  (16 POINT)"
 3000   PRINT "GIVES INTEGRAL OF";S
 3010   PRINT
 3020   SELECT PRINT 005
 3030   GOTO 1360
 3040   REM
 3050   END
```

investigatory purpose, and also to some degree outside our competence, but it does suggest that more serious work on the properties of different methods of numerical integration as applied to probability formula calculations is needed. A listing of the "BFTEST" program for one illustrative run is provided in Figure 3.4.

Before leaving this topic of numerical integration, it is important to mention one further idea which has emerged, but which has not been systematically pursued due to the limitations of the present project. This is simply the notion that when there are several joint parameters involved in a data-generating process, the logic of the Bayesian analysis method implies that one should seek to evaluate the marginal prior distribution of any one of those by integration across all of the others. This is an example of the need for n-dimensional integration, or repeated integration. Such computations, even when analytic, can be complex. If they must be approached numerically, they tax the powers of most computer systems now available.

This is an area in which there has been considerable development, mostly in the engineering and applied mathematics fields. Yet very little is known about the applicability of some of the procedures that have been developed there to the assessment and evaluation of prior belief distributions. For example, one general approach which has been worked on for some years is that of a stratified sampling of the points at which the integral is evaluated; another is that of adapting the panel size dynamically during the integration so as to preserve as much as possible of the important

Figure 3.4

```
76/11/07                    BFCHECK                  PAGE 1
PROGRAM LISTING

 1010  REM "BFCHECK"    14 NOV 75  2200
      :DIM A$64
 1020  INPUT "LABEL",A$
 1030  REM SEE BOX AND TIAO, 1973, PP. 104-109 FOR FORMULAS
 1040  M1=08
      :N2=08
      :FOR T8=0  TO 45  STEP 15
 1050  SELECT PRINT 01D(90)
      :PRINT HEX( 0C000000000000000000)
 1060  PRINT HEX( 0000000000000000000A);A$
      :PRINT HEX( 0000000000)
 1070  PRINT "CHECKS BEHRENS DENSITY AND CUMULATIVE PROBABILITY"
 1080  PRINT HEX( 000000)
 1090  PRINT "(USES SIMPSON'S RULE FOR INNER; SIMPSON'S OUTER)"
 1100  SELECT D
      :REM SELECT DEGREE MODE
 1110  REM INPUT "N1,N2,PHI",N1,N2,T8
 1120  T9=1+(TAN( T8))↑2
 1130  V1=N1*(TAN( T8))↑2/T9
      :V2=N2/T9
 1140  F1=N1-1
      :F2=N2-1
 1150  B9=(V2/N2)/(V1/N1+V2/N2)
      :B8=1-B9
 1160  B7=((F2/(F2-2))*B9)+((F1/(F1-2))*B8)
 1170  B6=((F2↑2)/((F2-2)↑2*(F2-4)))*B9↑2
 1180  B5=((F1↑2)/((F1-2)↑2*(F1-4)))*B8↑2
 1190  B4=B6+B5
      :B3=(V1/N1)+(V2/N2)
      :B2=4+((B7↑2)/B4)
 1200  B1=((B2-2)/B2)*B7
      :B1=B1*B3
 1210  D=B2
      :M=M2-M1
      :F3=B1
 1220  PRINT "BEHRENS SCALE FACTOR=";SQR( V1/N1+V2/N2)
 1230  PRINT HEX( 0000);"BEHRENS"," N"," M"," V"
 1240  PRINT HEX( 000000);"GRP 1",N1,M1,V1
 1250  PRINT HEX( 000000);"GRP 2",N2,M2,V2
 1260  PRINT HEX( 000000);"PHI=";T8;"  TAN(PHI)=";TAN( T8)
 1270  REM END OF B-F PREPARATION
 1280  REM
 1290  F4=(INT( 100000*SQR( F3)+.5))/100000
      :S=SQR( F3)
      :PRINT
 1300  PRINT "-T- D.F. =";D;"; -T- MEAN =";M;";  -T- SCALE FACTOR =";F4
 1310  REM CALCULATE K1, CONSTANT USED IN T APPROXIMATION
 1320  G9=(D+1)/2
      :GOSUB 1970
      :D3=G0
      :G9=D/2
```

Figure 3.4(continued)

```
76/11/07                          BFCHECK                        PAGE 2
PROGRAM LISTING

          :GOSUB 1970
          :D4=G0
   1330   K1=D3-(D4+LOG( S*SQR( #PI *D)))
   1340   REM K1=LN(CONSTANT) IS NOW READY
   1350   REM
   1360   E8=-(F1+1)/2
          :E9=-(F2+1)/2
          :REM E8 & E9 ARE EXPONENTS IN FNP
   1370   C1=COS( T8)
          :S1=SIN( T8)
          :REM C1 & S1 ARE CONSTANTS IN FNP
   1380   GOSUB 1730
          :REM EVALUATES CONSTANT K2 FOR B-F DENSITY VALUE
   1390   PRINT
          :PRINT "CONSTANT K1 = ";EXP( K1)
   1400   PRINT HEX( 000000):"CONSTANT K2 = ";K2
   1410   REM
   1420   L0=-10
          :U0=+00
          :H0=.10
          :L=-11.5
          :U=+11.5
          :H=.20
          :H9=H/3
          :H8=H0/3
   1430   REM
   1440   S5=(U-L)/H-1
          :S4=(U-L)/H-2
          :S3=(U-L)/H
   1450   PRINT
          :PRINT "D (OUTER) BOUNDS = ";L0;"  ";U0;"STEP   ";H0
   1460   PRINT HEX( 000000):"Z (INNER) BOUNDS = ";L;"  ";U;"STEP    ";H
   1470   PRINT
   1480   PRINT "   X    B DENSITY   B CUM   T DENSITY   T CUM    D DIFF     C DIF
          F      D %    C %"
   1490   J3=L0
          :GOSUB 1830
          :P2=P1*H8
          :REM GETS INITIAL DENSITY IN P2
   1500   PRINT
   1510   J3=J3+H0
   1520   GOSUB 1830
          :REM GETS DENSITY IN (P1*H8) VIA SIMPSON'S RULE
   1530   P2=P2+4*P1*H8
   1540   J3=J3+H0
   1550   GOSUB 1830
   1560   P2=P2+P1*H8
   1570   X=J3
   1580   GOSUB 2070
   1590   J4=FN T(X)
   1600   F2=5E-3
```

Figure 3.4(continued)

```
76/11/07                    BFCHECK                          PAGE 3
PROGRAM LISTING

      :R5=5E-6
      :R6=5E-7
      :R8=5E-9
 1610  PRINTUSING 1620  ,J3,P1+R7,P2+R5,J4+R7,C+R5,(J4-P1)+R7,(C-P2)+P6,100*(J
      4-P1)/P1+R2,100*(C-P2)/P2+R2
 1620  % -##.## #.####### #.#####  #.####### #.#####  -#.###### -#.###### -
      ##.## -##.##
 1630  P2=P2+P1*H8
 1640  IF J3<U0  THEN 1510
 1650  FOR I=1  TO 10
      :PRINT
      :NEXT I
      :PRINT HEX( 0C0000000000000000A)
 1660  P2=0
 1670  NEXT T8
      :PRINT HEX( 0C)
 1680  PRINT HEX( 0000000000000000000000000000000A)
 1690  STOP
 1700  REM
 1710  REM END OF MAIN PROGRAM: SUBROUTINES FOLLOW
 1720  REM ************************************************
 1730  REM SUBROUTINE TO GET CONSTANT K2, GIVEN F1,F2
 1740  G9=(F1/2)
      :GOSUB 1970
      :B1=EXP( G0)
 1750  G9=(F1+1)/2
      :GOSUB 1970
      :B2=EXP( G0)
 1760  B3=SQR( #PI )
 1770  G9=(F2/2)
      :GOSUB 1970
      :B4=EXP( G0)
 1780  G9=(F2+1)/2
      :GOSUB 1970
      :B5=EXP( G0)
 1790  B6=SQR( #PI )
 1800  K2=1/((B1*B3)/B2*(B4*B6)/B5*SQR( F1*F2))
 1810  RETURN
 1820  REM
 1830  REM SUBROUTINE TO GET DENSITY VIA SIMPSON'S RULE
 1840  K5=0
      :K6=0
 1850  W8=J3*S1
      :W9=J3*C1
      :K4=FN P(L)
 1860  FOR I=1  TO S5  STEP 2
      :K5=K5+FN P(L+I*H)
      :NEXT I
 1870  FOR I=2  TO S4  STEP 2
      :K6=K6+FN P(L+I*H)
      :NEXT I
```

Figure 3.4(continued)

```
76/11/07                        BFCHECK                      PAGE 4
PROGRAM LISTING

 1880   K7=FN P(U)
 1890   F1=K2*H9*(K4+4*K5+2*K6+K7)
 1900   RETURN
 1910   REM
 1920   DEFFN P(Z)=(1+(Z*C1-W8)*(Z*C1-W8)/F1)↑E8*(1+(Z*S1+W9)*(Z*S1+W9)/F2)↑E9
 1930   REM
 1940   REM
 1950   DEFFN T(X)=EXP( K1+LOG( (1+(((X-M)↑2)*(D+1))/(D*F3*(D+1)))↑(-(D+1)/2))
 1960   REM
 1970   REM THIS SUBROUTINE SETS G0=LN(GAMMA(G9))
 1980    G8=G9
        :IF G9<=1E30  THEN 1990
        :G0=1E38
        :RETURN
 1990    IF G9>1E-9  THEN 2000
        :G0=0
        :RETURN
 2000    IF G9<1E10  THEN 2010
        :G0=G9*(LOG( G9)-1)
        :RETURN
 2010    G7=1
 2020    IF 18<G8  THEN 2030
        :G7=G7*G8
        :G8=G8+1
        :GOTO 2020
 2030    G6=1/G8↑2
        :G0=(G8-.5)*LOG( G8)-G8+.918938533-LOG( G7)
 2040    G5=.833333333E-1
        :G4=.277777778E-2
        :G3=.793650794E-3
        :G2=.595238095E-3
 2050    G0=G0+1/G8*(G5-(G6*(G4+(G6*(G3-(G6*(G2))))))
        :RETURN
 2060    REM
 2070    REM SUBROUTINE TO GET T CUMUL PROB USING BETA DIST
 2080    REM INPUTS X,M,S,D; OUTPUT C
 2090    H1=1
 2100    T9=.0001
 2110    D7=-88
 2120    D8=+88
 2130    D9=100
        :REM MAX TERMS OF SOLUTION, REGARDLESS OF T9
 2140    D2=.5*D
 2150    A0=-1
 2160    T=(X-M)/S
 2170    Z2=.5*(1+T/SQR( D+T*T))
 2180    Z=Z2
        :IF Z2<.5  THEN 2190
        :Z=1-Z2
 2190    REM
 2200    IF Z<0  THEN 2790
```

Figure 3.4(continued)

```
76/11/07                    BFCHECK                    PAGE 5
PROGRAM LISTING

 2210   IF Z>1  THEN 2790
 2220   IF D2<0  THEN 2790
 2230   IF A0=D2  THEN 2330
 2240   G9=D2
 2250   GOSUB 1970
 2260   D3=G0
 2270   G9=D
 2280   GOSUB 1970
 2290   D4=G0
 2300   A1=2*D3-D4
 2310   A2=LOG( D)
 2320   A0=D2
 2330   A3=LOG( Z)
 2340   A4=LOG( 1-Z)
 2350   A5=(D2-1)*(A3+A4)-A1
 2360   IF A5<D7  THEN 2790
 2370   IF A5>D8  THEN 2790
 2380   H2=EXP( A5)
 2390   IF H1<>0  THEN 2410
 2400   RETURN
 2410   REM INITIALIZE FOR CUMUL
 2420   C=0
 2430   A6=Z/(1-Z)
 2440   A7=H2*(1-Z)/D2
 2450   REM EVALUATE BINOMIAL SERIES
 2460   E1=INT( D2)
 2470   IF E1<1  THEN 2620
 2480   E2=D9
 2490   IF E1>D9  THEN 2510
 2500   E2=E1
 2510   A8=0
 2520   FOR I=1  TO E2
 2530   E3=(D2-A8)/(D2+A8)*A6
 2540   IF A7>T9  THEN 2560
 2550   IF A7<(T9/E3-T9)  THEN 2760
 2560   A7=E3*A7
 2570   C=C+A7
 2580   A8=A8+1
 2590   NEXT I
 2600   IF E2<E1  THEN 2750
 2610   REM EVALUATE POWER SERIES
 2620   E4=E1
 2630   IF E4=D2  THEN 2760
 2640   A7=A7*(D2-E4)/(D2+E4)*Z/(1-Z)↑(D2-E4)
 2650   C=C+A7
 2660   E5=T9/A6
 2670   A8=E4+1
 2680   E6=E1+1
 2690   FOR I=E6  TO D9
 2700   A7=(D2+A8-1)/(D2+A8)*(A8-D2)/(A8-E4)*Z*A7
 2710   C=C+A7
 2720   IF A7<E5  THEN 2760
```

Figure 3.4(continued)

```
76/11/07                    BFCHECK                    PAGE 6
PROGRAM LISTING

 2730  A6=A6+1
 2740  NEXT I
 2750  PRINT "ITERATION DID NOT CONVERGE: CUM PROB MAY BE INACCURATE"
 2760  IF Z2<.5  THEN 2770
      :C=1-C
 2770  RETURN
      :REM THIS IS NORMAL RETURN: C = CUMUL PROBAB
 2780  REM HANDLE ILLEGAL AND SPECIAL CASES
 2790  PRINT "SPECIAL CASE ROUTINE IN USE"
 2800  A5=0
 2810  Z2=Z
 2820  IF Z<.5  THEN 2860
 2830  Z2=1-Z
 2840  GOSUB 2890
 2850  RETURN
 2860  GOSUB 2890
 2870  C=1-C
 2880  RETURN
 2890  REM SPECIAL CASES SUBROUTINE
 2900  C=0
 2910  H2=0
 2920  IF A5<0  THEN 2950
 2930  IF A5=0  THEN 2970
 2940  IF A5>0  THEN 2980
 2950  IF Z2>D2  THEN 2960
      :RETURN
 2960  C=1
 2970  RETURN
 2980  PRINT "ARGUMENTS BEYOND CAPACITY: SPURIOUS PROBABILITY"
 2990  H2=2E22
 3000  RETURN
 3010  END
```

fluctuation in the function. These two concepts have recently been combined in a single algorithm (Gallaher, 1973), but there has been no attempt to apply that logic, or any variant of it, to the task of actually making multidimensional parameter assessments in statistical analysis.

Here, as with the comparison of methods, our work as part of this project has been far from thorough. Instead, we have felt the important tasks to be those of verifying the accuracy of the Patil approximation, of drawing attention to these areas in general, and of providing a basic tool (the INTEG program) which may be used in subsequent work to investigate the numerical analysis aspects of Bayesian inference more carefully.

3.6 Programming the Inverse Chi Distribution

As has been discussed earlier, if we are to describe our knowledge about a data generating process which follows the Normal distribution, we must take into account three elements of that knowledge: (1) the assertion (equivalently, the postulation) that the data generating process is indeed Normal in form; (2) an expression of our knowledge about the mean (mu) of the data generating process, which expression usually takes the form of a standardized conditional likelihood curve for the parameter mu; and (3) an expression of our knowledge about the standard deviation (sigma) of the data generating process, which expression usually takes the form of a standardized conditional likelihood curve for the parameter sigma. We have indicated how generally desirable it is if the SCLs can be chosen from families of curves which are "conjugate" to the Normal distribution data generating process. If our knowledge can indeed be represented faithfully, or at least adequately, by a curve which is a member of the conjugate family, then two advantages are immediately obtained. First, the knowledge can be interpreted as equivalent to that from a specific prior sample. Second, the combining of the prior SCL with the information from the data is much simpler, and can be done recursively, so that cumulation of knowledge becomes almost unavoidable.

We have noted that when the standard deviation is not known a priori, but must be estimated from the data, then the distribution for the parameter mu which is conjugate is the Student's _t_. We have discussed at some length the evaluation of Student's _t_ curves, and the application of these ideas to comparisons of the two means when an experimental and a control group are measured. Thus far, we have _not_ given any attention to the appropriate form of SCL to be used for the other parameter of the Normal data generating process, the standard deviation, sigma. However, we have mentioned in passing the Normal/inverse chi model, and we now explain how this model fits into the larger picture.

Basically, the representation of our knowledge about the standard deviation sigma as following an inverse chi distribution is adopted so that this knowledge will be in a conjugate form. The rationale for the choice of this particular distribution is rather complex, and is discussed fully in Box and Tiao (1973) and in Novick and Jackson (1974). The distribution is, however, an important one, and we shall be using it in the later chapters.

Beyond the more fundamental issues as to the appropriateness of one or another distribution as the conjugate form, and the basic epistemological issues that are raised by the whole question of whether a conjugate form is warranted -- which we shall not discuss here -- there are also some incidental complications in dealing with the inverse chi. These complications are especially vexing to someone encountering this material for the first time, since they substantially retard absorption of the main ideas. One of these complications is that the chi-squared distribution and the inverse chi distribution are both special cases of the gamma distribution, and are described in that latter framework by some authors. For example, the MANECON manuals give various capabilities for the gamma distribution, but never mention the chi-square or inverse chi. If one was not aware of the relationships, and able to preprocess his input, he would believe the MANECON package simply neglects this problem. Other writers (e.g., Schmitt, 1969) likewise have their own terminology.

The second complication arises simply because the inverse chi, like most other distributions, itself has parameters which must be specified to completely identify the distribution. In particular, the parameters are the degrees of freedom (sometimes referred to as a shape parameter when the gamma distribution is being discussed), and the scale factor, a measure of spread. As with most of the basic distributions, this distribution is usually tabulated or referred to in its standardized form. That is, among the family of inverse chi distributions, those which have a scale factor equal to 1.0 are usually presented. In fact, the formula for the standardized version of the chi-square as well as that for the inverse chi are usually given as if no more general formulas are available. In this discussion, as with our previous treatment of the Student's t, we shall consider an unstandardized version of the distribution, one whose scale parameter is the basic way of relating it to the research information in mind.

We shall first give a brief discussion of the density formula for the inverse chi, drawing upon Box and Tiao (1973) and upon Novick and Jackson (1974). Consider a variable sigma (denoted σ), which we believe has an inverse chi distribution. Assume that we have measured N cases, and have calculated the usual unbiased (not maximum likelihood) estimator of the variance. Thus, we have f = N - 1 degrees of freedom, and we have

$$s^2 = \frac{\Sigma(X_i - \overline{X})}{f} \quad .$$

Now, let the expression IC(σ: f, s^2) denote the particular inverse chi distribution of interest. The formula for the density of that distribution is as follows:

$$IC(\ \sigma: f, s^2) = \frac{2 * (f*s^2)^{(f/2)}}{\Gamma(f/2) * (2)^{(f/2)}} * \sigma^{-(f+1)} * \exp\left[(f*s^2) \ / \ (2*\sigma^2)\right]$$

Notice that in the expression for the inverse chi density, the quantity $(f*s^2)$ = the sum of the squared deviations from the sample mean. This quantity appears twice in the expression, and is referred to as the square of the scale factor for the inverse chi distribution. That is, the scale factor of the inverse chi distribution is $(f*s^2)^{1/2}$.

Most discussions of the inverse chi, including the Tables published by Isaacs et al. (1974) report the density and/or cumulative values for a standardized inverse chi, one whose scale factor is 1.0. This is not the most convenient way to use the inverse chi distribution in actual data analysis, however, and so we shall work primarily with the unstandardized version of the distribution. To be able to do this, however, requires that the user be able to specify the particular distribution of interest among the various inverse chi distributions that are possible, by giving the degrees of freedom and the scale parameter. But, since the scale parameter is not commonly encountered, we make use of some relationships described in Novick and Jackson (1974: 195). They point out that the mode of the inverse chi distribution (i.e., the most likely or most frequent value, or the value given as the maximum likelihood estimate of the quantity) equals a simple expression:

$$\text{mode} = \left[\frac{(f*s^2)}{(f+1)} \right]^{1/2}$$

Therefore, if we have available (either from our prior judgment or from some data) a modal estimate of sigma, we can use it to derive the scale factor, and thus can evaluate the particular inverse chi distribution that is conjugate and consistent with the information at hand.

There is a routine in the CADA package which provides some descriptive evaluation of an inverse chi distribution, but it has serious limitations for research use. In particular, it requires that the inverse chi scale factor be given as input (something which is made available in the CADA package, but is not generally available in other types of statistical package), and it provides no graphic output of the distribution, nor any comprehensive report on its values. As mentioned earlier, the MANECON package rather successfully conceals its ability to evaluate an inverse chi, though the ability is included in the programs.

Therefore, as a way of gaining further insight into this distribution, and perhaps of improving the situation for research analysis somewhat, we have investigated the programming aspects of the inverse chi, and have developed a new program for evaluating such distributions. This new program has been named "ICD", and was written as a companion to the TBF program described earlier. The next few paragraphs will describe this ICD program and its features.

Basically, as with the Student's *t* distribution, the evaluation of the inverse chi *density* is straightforward. It is the evaluation of the *cumulative* probability curve that presents some complications. The approach taken is, however, much simpler than was the case with the earlier program, simply because the inverse chi cumulative formula is a simple transformation of the chi-square cumulative formula, and that latter formula is widely programmed. The CADA routine is fairly typical, in that it uses a numerical approximation (a 24 point Gauss-Laguerre quadrature) for small degrees of freedom to evaluate a standardized chi square cumulative probability, and then carries out simple transformations to obtain the result for the inverse chi, and to unstandardize it. The two problems of constructing a program are therefore: (1) to make the correct transform from an unstandardized inverse chi variable to a standardized chi-square variable; and (2) to choose an accurate and efficient procedure for evaluating the cumulative distribution. The first problem is trivial, since it has already been done in the CADA package. If X1 is an unstandardized inverse chi variable with scale factor F, then the new variable $Y = (F^2)/(X1^2)$ is a standardized chi-square variable.

The second problem is slightly more complex. The CADA algorithm using the 24-point Gauss-Laguerre quadrature is quite accurate, but not particularly fast. Also, the routine switches to an approximation formula (the Wilson-Hilferty cubic approximation based on a Normal curve) at 30 degrees of freedom. This is perhaps a bit sooner than would be ideal for accuracy. An alternative algorithm which appears attractive for many purposes is a recurrence relation embodied in Algorithm 299 of the C. A. C. M. series (Hill and Pike, 1967; Abramowitz and Stegun, 1968: 941, eq. 26.4.8). This latter algorithm is (for small degrees of freedom) quicker than the Gauss-Laguerre approach, and is perfectly accurate. However, it cannot handle fractional degrees of freedom, which may occur under some circumstances.

Therefore, the approach taken in writing the ICD program was to employ a composite of three different algorithms, and to use the one which was most suited to the actual problem. If the degrees of freedom are greater than a user-specified constant (presently set to default at 40), then the Wilson-Hilferty approximation via the Normal is used. If the degrees of freedom are small and not an integer, then the 24-point Gauss-Laguerre algorithm is used (with more accurate constants than those employed in the CADA routine). Finally, if the degrees of freedom are small and are integral, then the recurrence algorithm from Hill and Pike is used. This last provides a substantial increase in processing speed with no loss of accuracy for the cases where it can be used.

Also, the program offers similar output options as does TBF. That is, the output includes a list of values of the unknown quantity (sigma), with starting point, ending point, and step width specified by the user, and the density and cumulative probability at each of those points. It also provides a sketch graph of the density curve upon request, and with flexible control of the various output options. This program is currently operational in the language BASIC on the Wang 2200, and on the DECsystem-10. A copy of the listing is provided here as Figure 3.5. Notice too that an alternative available to the user is to couple this program with the output from a TBF run in which two means are being compared. Thus, the overall analysis is simplified somewhat, although the two programs also can be used quite independently.

Figure 3.5

```
76/11/20                         ICD                          PAGE 1
PROGRAM LISTING

 1010  REM    "ICD" DENSITY & CUMUL FOR INVERSE CHI DISTRIBUTION
 1020  REM
 1030  REM                     WRITTEN BY
 1040  REM                  JAMES FENNESSEY
 1050  REM            DEPARTMENT OF SOCIAL RELATIONS
 1060  REM            THE JOHNS HOPKINS UNIVERSITY
 1070  REM            BALTIMORE, MARYLAND  21218
 1080  REM
 1090  REM            20 NOVEMBER 1976  1730 VERSION
 1100  REM            (FOR USE WITH OUTPUT FROM DUAL -T- TBF)
 1110  REM
 1120  REM        THIS PROGRAM HAS BEEN CHECKED WITH REASONABLE
 1130  REM        CARE, BUT IS NOT GUARANTEED.  PLEASE NOTIFY THE
 1140  REM        AUTHOR OF ANY ERRORS.
 1150  REM
 1160  REM        THREE DIFFERENT ALGORITHMS ARE USED FOR THE
 1170  REM        CUMULATIVE PROBABILITY
 1180  REM        (A) IF D.F. > C7, USE WILSON-HILFERTY
 1190  REM            CUBIC-NORMAL FORMULA (PROCEEDINGS NAT. ACAD.
 1200  REM            SCI., VOL 17, 1931, PP. 684-688)
 1210  REM        (B) IF D.F. <= C7 AND INTEGER, USE ALGORITHM
 1220  REM            #299 (COMMUNICATIONS OF THE ASSOC. FOR COMP.
 1230  REM            MACH., VOL. 10, 1967, PP. 243-244)
 1240  REM        (C) IF D.F. <= C7 AND NON-INTEGER, USE ALGORITHM
 1250  REM            FROM 'NEW CDTR' ROUTINE IN THE "CADA"
 1260  REM            PACKAGE.  THAT ALGORITHM IS A 24 POINT
 1270  REM            GAUSS-LAGUERRE QUADRATURE.  (NOVICK AND
 1280  REM            ISAACS, UNIVERSITY OF IOWA, 1975)
 1290  REM
 1300  REM        THE NORMAL DISTRIBUTION EVALUATOR IS BASED ON
 1310  REM        EQ. 26.2.17 OF THE N.B.S. HANDBOOK (1968)
 1320  REM        OTHER RELEVANT EQS IN N.B.S. HANDBOOK ARE
 1330  REM        26.4.14 AND 26.4.6.
 1340  REM
 1350  REM
 1360  C1=1.330274429
      :C2=1.821255978
      :C3=1.781477937
 1370  C4=0.356563782
      :C5=0.319381530
      :C6=SQR( 2*#PI )
 1380  C7=40
      :REM C7 = CROSSOVER TO W-H FORMULA
 1390  DIM A$64
      :SELECT PRINT 005(64)
 1400  SELECT PRINT 005
 1410  PRINT HEX( 03);"INVERSE-CHI DISTRIBUTION EVALUATOR"
 1420  INPUT "TYPE 1 FOR OUTPUT ON PAPER; OTHERWISE 0",S9
 1430  PRINT "TYPE LABEL FOR PROBLEM (TO STOP, TYPE 'END')"
 1440  INPUT A$
 1450  IF A$="END"  THEN 1860
 1460  INPUT "TYPE 1 FOR DUAL T INPUT,OTHERWISE 0",S8
```

Figure 3.5(continued)

```
76/11/20                    ICD                         PAGE 2
PROGRAM LISTING

 1470  ON S8 GOTO 1530  , 1520
 1480  PRINT A$
 1490  IF S9<>1  THEN 1500
      :SELECT PRINT 015
      :PRINT
 1500  PRINT HEX( 03);"INVERSE-CHI DISTRIBUTION EVALUATOR"
 1510  GOSUB 2790
 1520  INPUT "D.F., UNBIASED ESTIMATE OF S.D.",F,S1
      :GOTO 1550
 1530  INPUT "N1, N2, DUAL T SCALE FACTOR",F1,F2,S1
 1540  F=F1+F2-2
      :S1=S1↑2
      :S1=S1/( 1/F1+1/F2 )
      :S1=SQR( S1)
 1550  S0=SQR( F*S1↑2)
      :REM S0 IS SQROOT OF (SUM OF SQUARED DEVS)
 1560  IF F>C7  THEN 1610
      :IF F<>INT( F)  THEN 1600
 1570  Q0=1
      :REM D.F. IS INTEGER <= C7
 1580  IF 2*INT( F/2)=F  THEN 1590
      :E1=1
      :GOTO 1630
 1590  E1=2
      :GOTO 1630
 1600  Q0=2
      :REM D.F. IS NON-INTEGER, <= C7
      :GOTO 1630
 1610  Q0=3
      :REM D.F. > C7
      :GOTO 1630
 1620  GOSUB 2790
 1630  INPUT "TYPE LOWER AND UPPER BOUND AND STEP WIDTH",L,H,W
 1640  IF L>0  THEN 1660
 1650  PRINT "X MUST BE > 0"
      :GOTO 1630
 1660  IF S9<>1  THEN 1670
      :SELECT PRINT 015
      :PRINT
 1670  PRINT A$
      :PRINT "D.F. = ";F
 1680  PRINT "MODE = ";S0/SQR( F+1);"   SCALE FACTOR = ";S0
 1690  PRINT " X-VALUE       DENSITY      CUMUL PROB"
 1700  G9=.5*F
      :GOSUB 2090
      :REM GETS LN GAMMA
      :T1=-G0
 1710  K=LOG( 2)-G0+(LOG( S0↑2/2))*G9
 1720  FOR X0=L  TO H  STEP W
 1730  X=S0↑2/X0↑2
      :REM CONVERTS TO STANDARDIZED CHI-SQR
 1740  C9=EXP( K+(-(F+1))*LOG( X0)-.5*X)
```

Figure 3.5(continued)

```
76/11/20                    ICD                              PAGE 3
PROGRAM LISTING

 1750  ON O0 GOTO 1760   , 1770 , 1780
 1760  GOSUB 1910
      :REM CHI299
      :GOTO 1790
 1770  GOSUB 2260
      :REM CHICHDA
      :GOTO 1790
 1780  GOSUB 2190
      :REM CHICUBIC
      :GOTO 1790
 1790  % ####.###    ####.####       #.####
 1800  PRINTUSING 1790   ,X0,C9+.00005,1-P+.00005
 1810  NEXT X0
 1820  PRINT
      :GOSUB 2790
      :PRINT
 1830  INPUT "DO YOU WANT THE PLOT",B$
      :IF STR( B$,1,1)<> "Y"  THEN 1360
 1840  GOSUB 2810
 1850  GOTO 1360
 1860  PRINT
      :PRINT
      :SELECT PRINT 005
      :STOP "END OF JOB"
 1870  REM OF MAIN PROGRAM
 1880  REM SUBROUTINES FOLLOW
 1890  REM
 1900  REM CHI-SQUARE ALGORITHM #299 FROM CACM
 1910  A=X/2
      :Y=0
 1920  IF E1=1  THEN 1930
      :GOTO 1940
 1930  IF F<=2  THEN 1950
 1940  Y=EXP( -A)
 1950  IF E1=1  THEN 1960
      :S=Y
      :Z=1
      :E=1
      :GOTO 1990
 1960  Y3=SQR( X)
 1970  D=1/C6*EXP( -.5*Y3↑2)
      :T=1/(1+.2316419*Y3)
 1980  S=2*D*T*((((C1*T-C2)*T+C3)*T-C4)*T+C5)
      :Z=.5
      :E=1/SQR( A*#PI )
 1990  IF F<=2  THEN 2050
 2000  X1=(F-1)/2
 2010  C=0
 2020  FOR I=Z  TO X1
      :E=E*A/I
      :C=C+E
```

Figure 3.5(continued)

```
76/11/20                          ICD                          PAGE 4
PROGRAM LISTING

      :NEXT I
 2030  P=1-(C*Y+D)
 2040  GOTO 2060
 2050  P=1-D
 2060  RETURN
 2070  REM
 2080  REM LOG GAMMA
 2090  IF G9<=1E30  THEN 2100
      :G0=1E38
      :RETURN
 2100  IF G9>1E-9  THEN 2110
      :G0=0
      :RETURN
 2110  IF G9<1E10  THEN 2120
      :G0=G9*(LOG( G9)-1)
      :RETURN
 2120  G8=G9
      :G7=1
 2130  IF G8>18  THEN 2140
      :G7=G7*G8
      :G8=G8+1
      :GOTO 2130
 2140  G6=1/G8↑2
      :G5=1/12
      :G4=1/360
      :G3=1/1260
      :G2=1/1680
      :G1=LOG( SQR( 2*#PI ))
 2150  G0=G1+(G8-.5)*LOG( G8)-G8+1/G8*(G5-(G6*(G4+(G6*(G3-G6*G2)))))-LOG( G7)
 2160  RETURN
 2170  REM
 2180  REM WILSON-HILFERTY APPROX VIA CUBIC & NORMAL
 2190  Y3=((X/F)↑(1/3)-1+2/(9*F))/SQR( 2/(9*F))
      :Y4=ABS( Y3)
 2200  D=1/C6*EXP( -.5*Y3↑2)
      :T=1/(1+.2316419*Y4)
 2210  P=1-D*T*((((C1*T-C2)*T+C3)*T-C4)*T+C5)
 2220  IF Y3>0  THEN 2230
      :P=1-P
 2230  RETURN
 2240  REM
 2250  REM CADA CHI-SQ ROUTINE
 2260  IF F>C7  THEN 2190
 2270  T0=.5*F-1
 2280  X=X/2
 2290  T=81.4982792
 2300  P=5.57534579E-35*(T+X)↑T0
 2310  T=69.9622400
 2320  P=P+4.08830159E-30*(T+X)↑T0
 2330  T=61.0585314
 2340  P=P+2.45181885E-26*(T+X)↑T0
```

Figure 3.5(continued)

```
76/11/20                    ICD                         PAGE 5
PROGRAM LISTING

 2350   T=53.6085745
 2360   P=P+3.60576586E-23*(T+X)↑T0
 2370   T=47.1531064
 2380   P=P+2.01051746E-20*(T+X)↑T0
 2390   T=41.4517205
 2400   P=P+5.35018881E-18*(T+X)↑T0
 2410   T=36.3584058
 2420   P=P+7.81980038E-16*(T+X)↑T0
 2430   T=31.7760414
 2440   P=P+6.69418105E-14*(T+X)↑T0
 2450   T=27.6359372
 2460   P=P+3.91773652E-12*(T+X)↑T0
 2470   T=23.8873298
 2480   P=P+1.50700823E-10*(T+X)↑T0
 2490   T=20.4914601
 2500   P=P+4.07285899E-09*(T+X)↑T0
 2510   T=17.4179926
 2520   P=P+7.96081296E-08*(T+X)↑T0
 2530   T=14.6427323
 2540   P=P+1.15131581E-06*(T+X)↑T0
 2550   T=12.1461027
 2560   P=P+1.25447220E-05*(T+X)↑T0
 2570   T=9.91209802
 2580   P=P+1.04461215E-04*(T+X)↑T0
 2590   T=7.92753925
 2600   P=P+6.72162564E-04*(T+X)↑T0
 2610   T=6.18153512
 2620   P=P+3.36934906E-03*(T+X)↑T0
 2630   T=4.66508370
 2640   P=P+.0132260194*(T+X)↑T0
 2650   T=3.37077426
 2660   P=P+4.07324782E-02*(T+X)↑T0
 2670   T=2.29256206
 2680   P=P+9.81662726E-02*(T+X)↑T0
 2690   T=1.42559759
 2700   P=P+.183322689*(T+X)↑T0
 2710   T=.766096906
 2720   P=P+.258806707*(T+X)↑T0
 2730   T=.311239146
 2740   P=P+.258774108*(T+X)↑T0
 2750   T=5.90198522E-02
 2760   P=P+.142811973*(T+X)↑T0
 2770   P=1-EXP( T1+LOG( P)-X)
 2780   RETURN
 2790   PRINT "---------------------------------------"
       :RETURN
 2800   REM PLOTS Y VERSUS X
 2810   INPUT "TYPE LOWER AND UPPER BOUND AND STEP WIDTH",L,H,W
 2820   W1=INT( (H-L)/W+1)
 2830   IF W1<=200  THEN 2840
       :PRINT "TOO MANY POINTS"
```

Figure 3.5(continued)

```
76/11/20                    ICD                           PAGE 6
PROGRAM LISTING

     :GOTO 2610
 2840  X0=S0/(SQR( F+1))
 2850  X=S0↑2/X0↑2
     :P=EXP( K+(-(F+1))*LOG( X0)-.5*X)
 2860  Y9=P/50
     :REM Y9=ONE PRINT POS
 2870  SELECT PRINT 015(80)
     :PRINT
     :GOSUB 2790
     :PRINT A$
 2880  PRINT "INV CHI DENSITY ON ";F;" D.F."
 2890  PRINT "MODE = ";X0;"   SCALE FACTOR = ";S0
 2900  Y8=INT( Y9*10↑4+.5)/10↑4
 2910  PRINT "X-AXIS IS * IF DENSITY <";Y8;"; OTHERWISE IS I"
 2920  PRINT TAB( 16);"TICK MARKS ON Y-AXIS EVERY";5*Y8
 2930  PRINT
     :PRINT TAB( 16);"0";TAB( 64);50*Y8
 2940  PRINT TAB( 16);"I----:----I----:----I----:----I----:----I----:----I----
     :"
 2950  FOR J=1  TO W1
     :B$="*"
     :C$="I"
     :X0=L+(J-1)*W
 2960  X=S0↑2/X0↑2
     :P=EXP( K+(-(F+1))*LOG( X0)-.5*X)
 2970  IF P>=Y8  THEN 2980
     :B$=" "
     :C$="*"
 2980  Y=INT( P/Y9+.5)
     :X0=INT( X0*10↑4+.5)/10↑4
 2990  PRINT J;TAB( 6);X0;TAB( 16);C$;TAB( 16+Y);B$
     :NEXT J
 3000  PRINT
     :GOSUB 2790
     :PRINT
     :SELECT PRINT 005(64)
     :RETURN
 3010  REM $
```

4.0 A Bayesian Reanalysis of the Incentives in Education Data

4.1 Introduction

One of the earliest areas in which applied social research has developed on a large scale, and yet also one in which a sense of little progress in endemic, is the study of alternative educational programs, at the various levels of the school, of the classroom, and of the teacher-student relationship. Since approximately 1964, this area of research has taken on a new cast, as a result of the social activism of the "Great Society" programs sponsored by the Federal government. A variety of political circumstances led to the emergence of a federal policy area known as compensatory education. The features of that social movement included a new level of active involvement in educational management by many actors--federal and local officials, community and parent groups, educators and university researchers. Numerous different approaches to the improvement of education of children who meet various definitions of "disadvantaged" were developed and implemented, with various parties acting as the sponsor, consumer, and/or critic of each kind of effort. Many of these efforts have been relatively short-lived; others are now an established part of educational practice.

One distinctive feature of the movement toward compensatory education is that its large-scale structure, as reflected in the Federal legislation of the ESEA of 1965, included a redefinition of the important inputs to any educational decision-making process. One aspect of this shift, already mentioned, was the specific and active involvement of the "local community" (parents and local interest groups) in developing and monitoring the operation of school programs. A second novelty was the specific requirement that empirical evaluation of the effects of each new program be done, using the best feasible scientific methods. In short, the compensatory education movement included not only a redefinition of the challenges confronting public education, but also (though less visibly) a redefinition of the legitimate procedures for making educational policy and program decisions. In two respects, the autonomy of professional educators was weakened: first, the movement encouraged the active representation of additional local interests in any decisions (community participation); second, the movement also encouraged the introduction of additional kinds of empirical information, the kind best collected by trained scientists. Our interest in this chapter is primarily in the latter feature.

The evaluation aspect of the compensatory education movement was in itself a major innovation in the creation and social use of social information. However, because it emerged as an apparently incidental part of another movement, the evaluation aspect was always regarded as less than highest priority. As a consequence, there has been considerable variety and fluctuation in the kinds of evaluation being done. The conceptualization, design, implementation, and execution of such studies has ranged from excellent to shameful.

The large majority of the studies are afflicted with serious flaws. The findings from such studies likewise have been varied in content, although the majority of them tend to show results considerably less successful than the program sponsors had hoped for.

Improving the value of these evaluation studies and developing their methodology in a broad sense thus stand as under-recognized problems for professional education and for applied social science. In this chapter, we concentrate on one such evaluation study as an illustration, and apply to it some Bayesian tools for analysis, inference, and presentation of results. The aim here is to show that the original report of that study, while adequate, was less useful than it might have been, and that the use of a different paradigm of analysis and inference makes possible a more effective interpretation. The particular study chosen as the vehicle for this discussion has some attributes which make it especially suitable. One such attribute is that the original data analysis for the project report was performed by this author, so that criticism now of the decisions and methods adopted then will be possible without the risk of engendering an irrelevant and personalized debate about the merits of the original analysis.

The study to be used here, known as the OE Incentives Study, was basically quite simple in design. Its aim was to provide an initial assessment of one straightforward though highly unorthodox alternative program arrangement for public elementary schools enrolling a large percentage of disadvantaged students. In other words, the conceptual definition of the "treatment" to be employed is uncomplicated. Also, the study was clearly a product of the larger ebbs and flows of the compensatory education movement, beginning at a time when that movement was just past its peak (1971). When the OE Incentives study was being planned, the education community had recently received and absorbed the major evaluation of Project Head Start (Westinghouse/Ohio University, 1969). Other smaller scale reports of particular compensatory programs had been published. Most of these reports contained disappointing numerical results and long explanations. Thus, the level of prevailing optimism and naivete was less than had been found several years earlier. A need for some new approaches to compensatory education was increasingly apparent. The OE Incentives study is of an interesting size and sponsorship. That is, it was neither extraordinarily large nor was it minor. Its budget, including costs of the treatment, was nearly $800,000, and it operated over a period of 15 months. The project was actively supported by the U.S. Office of Education and actively managed by two private research companies, with highly respected talent assigned to it. The analysis was done by a university-affiliated researcher, and was reviewed by some eminent experts in educational research prior to the drafting of the final report. Thus, it cannot be dismissed as an unfairly chosen example, nor is it so atypical in scale as to lack relevance to other studies. Its limited scale and short timetable (compared to the larger issues of concern) also exemplify the information-gathering problems confronting educational policy formation institutions.

The present reanalysis, then, is intended first as one contribution to a sociology of information processing, focusing on one instance of the social creation and use of formalized social knowledge. Second, it is an attempt to draw some further meanings for educators from a set of data which have been somewhat neglected, although they were collected at considerable cost.

4.2 The O.E. Incentives Study

In the spring of 1971, one of the "hottest" ideas among Federal specialists in education was the notion of "performance-contracting". This idea, compatible with the general political and ideological climate then prevailing, had originated with a small, privately sponsored effort in Texarkana, Texas, and had been gaining notoriety for several months. Basically, the idea behind this particular compensatory education strategy was borrowed from a practice best developed in the Defense Department, by which the contracting agency entered into an agreement with a private corporation. The agreement provided that the private corporation would develop a specified product to meet certain specified performance levels. Payment to the private contractor was contingent upon the demonstration that these standards had been achieved. The approach had been quite successful in reducing procurement management costs for a variety of technical products, since only the product itself needed close evaluation. In short, to use the educator's jargon, the performance contract concentrated evaluation effort on the result, the product, rather than on the process of its creation.

The adaptation of this procurement evaluation idea to compensatory education was fairly direct. A local education agency (i.e., school board) would enter into a legal contract with an organization (typically, a private, for-profit corporation), according to which the private organization would staff and operate an instructional program on specific topics for specific students. The key element of the contract was that this company would receive payment according to a formula based upon the scores achieved by its pupils on a specified battery of objective achievement tests. The idea possessed several features making it attractive to school boards increasingly eager for ways to deal with the challenge of obtaining higher achievement levels among their disadvantaged pupils. The Texarkana trial of this idea had apparently led to very sizeable gains in achievement in a short time, among a group of seriously disadvantaged students. The findings from that Texarkana study were later questioned seriously as perhaps the result of illegitimate coaching. However, they became public when the need for any success--even better, one which could be obtained at a reasonable cost and on a broad scale--was seen as nearly desperate. In a time of fiscal caution, the idea that the contract provided "no charge if we fail" also seemed an important evidence of financial astuteness by the school board.

The first Federal agency to react seriously to this idea was not the U.S. Office of Education, but the Office of Economic Opportunity. Appealing to its general mandate to assess ways of improving the life-chances of poor persons, this agency quickly mounted a large study of the performance contracting idea. That study began in the summer of 1970 and collected data during the 1970-71 school year. Thus, the U.S. Office of Education found itself in the role of spectator, at least temporarily.

Naturally, not all actors on the education scene found performance contracting attractive. Two groups which tended to be opposed to it were teacher unions and associations, and many parent representatives. They saw this approach as removing from them a sizeable portion of the control they had exercised over the content of school activities under more traditional school operating styles. These and other reactions to performance contracting led to the creation of a related but distinct idea as a strategy for compensatory education, one which was developed rapidly and actively within the U.S. Office of Education. This second proposed innovation was simple but elegant: apply to parents and to teachers the motivational stimulus exemplified in the performance contract--namely, payment increments for demonstrated results. This new idea thus hybridized the several ideas behind performance contracting, in a mainly pragmatic fashion. The teachers and parents retained their position of influence over school activities, but were now concretely stimulated with tangible rewards contingent upon the achievement level of their students. Fundamentally, the idea was the same as that of bonus payments to workers in many industries, but it embodied several distinctive features reflecting the educational situation.

As the idea took shape, several events were set in motion. The U.S. Office of Education commissioned a feasibility study (done by

A.I.R.) to ascertain whether a field study was advisable, and if so to suggest designs for one. Not surprisingly, the feasibility study concluded that a program of cash incentives to teachers and to parents appeared generally feasible, although some potential obstacles were identified. The same report suggested in broad outline a medium-scale field test design, involving several thousand students in a handful of schools. Within a few months of the submission of this report, the U.S. Office of Education issued an R.F.P. for the Incentives Study in the spring of 1971. In June 1971 two contracts were awarded for the execution of the study, one for the technical data collection and analysis, and a separate contract for a more verbally oriented management analysis of the incentives program.

The general design of the Incentives Study is in many respects quite sound and desirable. Its experimental design was simple and straightforward. Its plan for data collection efforts was rich and highly varied. The intention was to use what amounted to a randomized blocks design by choosing matched pairs of schools within each of several sites, and then randomly picking one school from each pair as a treatment school, the other serving as a control school. There was to be a good blend of quantitative data and qualitative data, plus allowance for substantial on-site support and monitoring.

The two treatments chosen for detailed investigation were: (1) payment of cash bonuses to teachers based on the average gains in reading and in math made by their students; (2) payments of cash bonuses to teachers and to parents, again based on the average gains in reading and in math made by their students. Each of these was to be contrasted with a control condition in which no cash incentives were offered. The fourth logical possibility in this design situation, a scheme of cash incentives to parents only, was not tried. Such a scheme was felt to be too divisive and threatening to the teachers. Omitting the fourth cell reduced the amount of data collection costs for the study, but also led to the necessity for only very direct analysis comparisons.

Because it was viewed as a "first cut", somewhat preliminary in nature, the study was designed with only two replications of each treatment. A total of four sites (two per treatment) and eight schools (EXP and CON in each site) were to be involved. The study officially began in June 1971, with the first wave of data collection scheduled for September 1971.

As events unfolded, a variety of problems, some foreseeable and others less obvious, occurred. Administrative and technical delays were encountered in tooling up for the data collection. Thus, although the achievement test data were collected very close to the intended schedule, much of the additional data collection on attitudes and behaviors of students, teachers, and parents was postponed, sometimes for as long as three months. Similarly, the original intent to randomly pick the treatment and the control school within each site from a pair of schools matched as closely as possible likewise foundered on the rocks of local administrative reality. As might be expected, there was pressure in each site that the incentive treatment be given to one school rather than the other, and this pressure generally did determine the assignment of schools to treatment and control conditions. Thus, the purity of random assignment, so essential to true experimental design, was lost. Finally, resistance to the philosophy of the experiment also was encountered. Many teachers strongly opposed the idea that their work should be placed on a commission basis; they expressed serious doubt that the cash incentives would produce any benefits, and in many cases they actively refused to cooperate. Various delays and communication problems in the Washington government offices also caused the implementation to be less decisive than had been intended. Another issue arose concerning the "fairness" of the formula originally proposed for determining payment. After some discussion in Washington and some expressed mistrust from the local sites, a revised formula was developed, which in large part met the objections. However, its implementation complicated the analysis and used up scarce time and money.

In April and May 1972, the post-test wave of data was collected. Again, the achievement data were collected according to plan, but the attitudinal and behavioral data were seriously incomplete. The analysis was done during the summer of 1972. Most of the analysis time was spent working up the attitude and behavior data, which had been collected at considerable cost. (This aspect of the project involved on-site research personnel in each city, repeated interviews with each teacher and many parents, and informal "field-work" observation reports.) The attitude and behavior aspects of the study had been intended to provide material which would suggest the "why"; the achievement test results were expected to tell "whether".

The analysis of the achievement test data was done via two approaches, one a simple comparison of gains, and the other a procedure which attempted to avoid problems of "fan-spread"--the differential gains typically made by different students depending upon their initial starting position. Since the experimental schools had lower starting scores than the control schools in most of the comparisons, this problem was a real one. Also, because of the way the treatment design had been developed, and given the research questions of main interest, the decision was made to use a Student's _t_ test statistic for the pairwise comparison of the relevant means, rather than to attempt a more complex analysis of variance formulation. The basic results thus were simply a comparison of gain scores for students under different treatments, adjusted for fan-spread. Despite the many technical details, important in themselves, the basic logic of the study was entirely adequately served by this analysis.

The findings were _not_ clear and simple. First, for reading there appeared to be no meaningful difference when the teachers-only treatment was compared to its control. This was also true for math. For the treatment cell using payments to parents and teachers, there was a modest favorable effect on reading. For the same treatment, however, the effect on math scores was quite large. The problem was that this pattern of findings, apparent when the data were aggregated to the school level, included many inconsistent reversals within school between grades, and within treatment between sites. No plausible explanation for these instabilities was apparent.

The original report of the project made several reasonable and cautious conclusions, and some sensible if perhaps optimistic recommendations as to the kinds of research procedure needed. It was received with thanks by the U.S. Office of Education, and thereafter was generally neglected.

4.3 Reanalysis from a Bayesian Perspective

As indicated in Chapter 2, the essence of quantitative social research, sometimes not fully appreciated by its practitioners or clients, is the specification of a generalized formal model for the data-generating process. This is followed by either the summarized description of the data via parameters of the model, or the comparison of alternative versions of that model with the actual data, to ascertain the relative and absolute goodness of fit of each version. The branch of mathematics called probability theory studies the abstract behavior of such models, taking as given their form and their parameters. The discipline of mathematical statistics, conversely, although it likewise takes the abstract model as given, studies the inferences possible about the parameters of the process, taking the data as given. It thus serves the practicing scientist by helping him to estimate the "constants" in the scientific laws.

For a Bayesian, the work of an empirical investigation begins, as it does for the classicist, with the choice of a particular family of data-generating models as being reasonable representations of the actual process. Such choices are made as considered professional judgments by the responsible investigator. If they are successfully challenged, then the entire analysis must be done again, using a new family of models. Although these choices are thus crucial, and there is much that can be said about the ways they might be improved, they are not the topic here. In fact, they are not approached much differently by the classicist or by the Bayesian, except perhaps that the Bayesian tends to be more conscious of the details of the data generating process model, since he uses it more directly and in more detail in his analysis.

In reanalyzing the O.E. Incentives study, the focal outcome variable will have to be drawn in some way from the scores achieved by individual students in testings at the beginning and at the end of the academic year, for total reading and total mathematics, using the Iowa Tests of Basic Skills. The "model" to be specified for this data thus must describe a process which can generate such scores. In other words, for each student the model must yield a score which has some systematic and some "random" components. In the original analysis and in this one, it has been deemed reasonable to adopt a linear model with the assumptions that the disturbance terms are normally distributed. Such a choice involves several tradeoffs. First, one must choose, on substantive grounds, a form of model (family of possible processes) which seems closest to the actual processes. This must be balanced against analytic manageability. In any case, the process of model formalization unavoidably implies some restriction of possibilities as compared with the verbal descriptive statements from which it is derived (cf. Land 1971).

The model specification, in other words, consisted of an interlocked series of decisions. These decisions were made by the investigator mainly in private, and somewhat unsystematically (cf. Cyert and March, 1963; March and Simon, 1958). We shall explicate them here as part of the overall attempt to elucidate the importance of judgment in the research analysis--even when the analysis is being done in the classical tradition.

Since the design of the O.E. Incentives study had attempted to equate the schools used as treatment and control as nearly as possible, it seemed highly likely that there would be no major differences in starting level of score, on the average, between the EXP school and its matching CON school. At the same time, there were reasons to expect some differences, so that simply comparing posttest scores of the two schools would be somewhat imprecise. Also, the fact that randomization in the treatment assignment had not in fact been achieved meant that there was probably some bias. Note that the regression of schools within site to a common mean is not likely if the schools really are different. So, the failure of randomization, if it is complete, at least removes regression to the mean (in one sense) as a threat. A straightforward approach to deal with this kind of design is to measure the average gain observed in each school. This approach has an additional advantage, since the gain was the criterion used for determining payments to the teachers and parents in the several EXP schools. Thus, our analysis would be close to the definition of the outcome variable actually being operated on by the treatment.

At this point in the logic, however, a complication arose. The comparison of change scores between EXP and CON schools might also be subject to the distortion of "fan-spread", a tendency for students with higher initial scores to gain more rapidly (i.e., to have gain rates approximately proportional to their scores) Through some fairly detailed logic which is spelled out in the original Incentives Project report (Planar Corporation, 1972), this consideration led to the additional use of an equivalent comparison. Instead of comparing only the observed gains in the original test score metric, another formal comparison was made of the Student's t statistic between EXP and CON before the treatment, with the corresponding t after the treatment. In addition, tabular and graphic results showing the gains in EXP and CON schools in the original score units also were presented in the original Incentives Project report.

One point on which the original analysis decisions were difficult was in defining the analytic units to be considered as the basis of comparison. Quite honestly, although we did contemplate some of these issues in 1972, we could not see how, in a classical statistics tradition, to make useful statements if we regarded the data as comprising only, say, 4 or 8 cases. The sample size available simply would be too small. This consideration led to a concentration only on the individual students; this choice was basically made for a combination of valid reasons and reasons of expediency. One factor in the decision was that the problem focus for the research had never been operationally specified in the project RFP. Yet, it was clearly a policy-research study rather than a purely disciplinary one. Therefore, I did not feel that I had the autonomy to make a personal decision that would be unorthodox and that was clearly contrary to the spirit of the design outlined in the RFP. I chose therefore to conduct the analysis assuming the relevant units (cases) were the individual students. This provided me with a "large sample" situation--approximately 3600 cases. However, it also traded away precision for simplicity. This still seems acceptable, but it neglected the real issue of school-to-school fluctuation.

That choice was reasonable and defensible. By making it, we were implicitly saying that our primary focus of interest was on disadvantaged students. However, it was not the only possible choice on strictly analytic grounds. The other possibilities were left unexamined in large degree because it was clear that the conventional analysis approaches using schools as the cases would make the study appear to be of little use, since a case base of only a few cases would be presented. The Bayesian approach, however, encourages such presentation. Note that this analytical judgment has nothing whatever to do with the assessment of a prior belief distribution. We are talking here only about an aspect of the appropriate data generating process to be studied.

Interestingly, if we assume that the data-generating process for the individual student scores is Normal, then it follows that any linear combination of such scores is also Normal. This implies that change scores (gain scores) on individual students, and also gain scores for classrooms, for grades, or for schools, all are Normal. In other words, the same abstract family of processes is implied whichever level of aggregation we decide to study, if the postulate is adopted that the individual scores occur with a Gaussian stochastic component.

This observation does simplify things in some respects. However, it leaves us still with the choice of the proper level of aggregation. There are several aspects to this choice. Most simply, there are four reasonable possibilities for aggregation levels. First, one might analyze the individual students, as was done in the original report. They are the units on which the main outcome measure was collected. Second, one might analyze teachers as the units. They are the direct recipients of the payment, and so are the "units" at the most disaggregated level to which the experimental treatment is being applied. Third, one might analyze results aggregated to the grade level. This could be viewed as a necessary substitute for working with the individual teachers as units, since the assignment of students to teachers is complex. Finally, and most reasonable as a policy input from the Federal government point of view, one might simply use schools as the unit of analysis. This would reflect the fact that the proposed educational policy under study would never be implemented except in an entire school. Thus, despite the content of the RFP, the focus of concern for the Federal policy decision--at least the most fundamental question--is whether the treatment, applied to a school, creates overall a dependable improvement in the achievement gains in that school.

To use individual students as the units implies a total case base of about 3600 (1800 per treatment). To use individual teachers implies a case base of about 160 (80 per treatment). To use grades implies a case base of 24 (12 per treatment). Finally, to use schools implies a case base of 8 (4 per treatment).

The choice of higher or lower level of unit involves two distinct kinds of considerations. First, if a higher level of aggregation is used, it is clear that information on the amount of variation in outcomes at the next lower level is of interest for policy formation as well. The Federal policy formation decision may be operative at the school level, but it is still of interest to Federal administrators to know whether the policy outcomes can be expected to be consistent or diverse within a single school. Second, to use a disaggregated level and to treat it for purposes of significance testing or confidence intervals as if each of the observations is "independent" clearly would not be consistent with the real-world process. That is, the students within a single classroom are not exposed to the treatment independently of each other; instead, they constitute a social unit, with many mutual influences among them. A similar point holds for the teachers in a school. These teachers work together, and discuss their reactions and experiences. The outcomes for one teacher are thus not statistically independent of those for another teacher in the same school. These facts imply that the choice of degrees of freedom for any statistical analysis is somewhat moot. The degrees of freedom is clearly fewer than the number of cases, but there is no feasible way of deciding how much fewer, at least not with this kind of data.

Because these considerations make clear that no single level of analysis is fully appropriate to the different aspects of the research questions under study, we have elected in this reanalysis to consider separately the results at the individual level (as in the original report), at the grade level, and at the school level. The reanalysis at the teacher level has been omitted for a combination of technical and practical reasons. The major technical reason is that many of the students were ability-grouped by teacher, so that there is an artifactual aspect to the levels of scores associated with many of the teachers. Also, many of the teachers had changing enrollments during the eight month period of the study, making the logical problems of assigning students to classes quite difficult. In addition to these logical problems, there was the practical consideration that any attempt to perform such an analysis would have required detailed work with the original data tapes, a task beyond the scope or need of this present investigation.

4.4 Reanalysis at the Level of Individual Students

The major results from the original analysis are presented in Tables III-3 and III-4 of the original Incentives Project Report

(p. III-16). These two tables show the averages of the gains achieved by individual students, in the EXP and in the CON schools, for reading and mathematics, respectively. Several different comparisons of EXP students versus CON students are presented in those tables. The most important of the comparisons made in these original tables are reproduced here as Tables 4.1 and 4.2.

Notice that Table 4.1 presents the mean, standard deviation, and number of cases separately for the EXP and the CON students, for each of several comparisons. The test score format used in these comparisons is a specially derived score provided by the publisher, and labeled as standard scores. They are referred to in this discussion as HBJ (Harcourt, Brace, Jovanovitch) standard scores. In the original analysis, for each score comparison, a Student's _t_ was computed. In addition, a point-biserial correlation between gain and treatment was computed, and serves as a measure of proportion of variation explained by the treatment. Also, a special statistic, developed for the O.E. Incentives Study, and labeled "Percent Net Impact", was calculated. This statistic reports the impact (gain in the EXP group) as a proportion or percentage of the CON group gain; it is thus analogous to a percentage markup in business. It has been provided to allow the reader to obtain a substantively meaningful interpretation of the magnitude of the observed impacts.

Consider in some detail the first results shown in Table 4.1. The top row of that table shows the results obtained for all students in the "TO" or "_t_eacher-_o_nly" segment of the Incentives Study. As

TABLE 4.1

TOTAL READING GAIN

COMPARISONS	GAINS						STATISTICS		
	EXPERIMENTAL GROUP			CONTROL GROUP			t-val.	r_{pb}	% Net Impact E/C
	$\overline{X}$ Gain	S.D.	N	$\overline{X}$ Gain	S.D.	N			
1. E/C, TO	6.48	5.81	1147	6.42	5.69	1055	+0.22	+.000	+ 1%
2. E/C, PT	7.59	6.64	694	6.01	6.28	775	+4.69	+.122	+26%
3. E/C, All	6.90	6.16	1841	6.25	5.95	1830	+3.25	+.054	+10%

TABLE 4.2

TOTAL MATHEMATICS GAIN

	GAINS						STATISTICS		
COMPARISONS	EXPERIMENTAL GROUP			CONTROL GROUP			t-val.	r_{pb}	% Net Impact E/C
	$\bar{X}$ Gain	S.D.	N	$\bar{X}$ Gain	S.D.	N			
1. E/C, TO	9.38	7.81	1144	8.79	7.62	1037	+1.81	+.039	+ 7%
2. E/C, PT	11.29	8.96	691	7.51	7.94	775	+8.56	+.218	+50%
3. E/C, All	10.10	8.31	1835	8.24	7.78	1812	+6.98	+.115	+23%

indicated in Table 4.1, among all students in the CON schools for this treatment, the mean gain in reading standard (HBJ) score was 6.42 points. The standard deviation of those gains was 5.69, and there were 1055 students in that category. Correspondingly, for the EXP students in this segment of the study, the mean gain was 6.48, the standard deviation was 5.81, and there were 1147 cases.

Using these results as the input to a conventional Student's _t_ analysis leads to a _t_ value of 0.22, as is shown in the tables reproduced in 4.1. There is essentially no variation accounted for by the treatment, and the percentage markup statistic is only 1%. In short, the TO treatment appears ineffective on reading gains.

We turn now to a Bayesian interpretation of these same results. This Bayesian reanalysis was performed first using a so-called "indifference" prior distribution. That is, a prior distribution was chosen to represent that particular pattern of prior knowledge which would have the _least possible_ impact on our final knowledge. In short, the prior approximates a state of ignorance. The Bayesian results corresponding to Table 4.1 are given in Table 4.3; the results corresponding to Table 4.2 are given in Table 4.4.

The Bayesian results using the indifference prior are quite similar to the original in actual content. This fact should come as no surprise, since both analyses are based on the same data. However, the Bayesian results are presented not in terms of a single _t_ value, but rather in terms of a complete distribution of Student's _t_. That complete Student's _t_ distribution formally represents our complete

TABLE 4.3

READING STANDARD SCORE GAINS, INDIVIDUAL LEVEL ANALYSIS
TEACHER PAYMENT ONLY, CONTROL VERSUS EXPERIMENTAL

	N	Mean	Variance
Control	1055	6.42	32.3761
Experimental	1147	6.48	33.7561

Student's _t_ distribution for Impact:

degrees of freedom = 2200 Mean = 0.06 Scale = .2454

(1) Impact Value	(2) Density	(3) Cumulative Probability
-1.0	0.0001	0.0000
-0.5	0.1205	0.0113
0.0	1.5776	0.4034
0.5	0.3259	0.9635
1.0	0.0011	1.0000
1.5	0.0000	1.0000
2.0	0.0000	1.0000
2.5	0.0000	1.0000
3.0	0.0000	1.0000

knowledge, posterior to the analysis of the study data, about the following quantity: the mean individual impact of the TO treatment on reading gain. Notice that we are conceptualizing the comparison in terms of the impacts upon individuals, even though we do not have direct information about the gains made by individuals. The general analysis model allows us to infer that the "mean of the differences" will equal the difference of the means. This is equivalent to saying that we can use the difference of the means as evidence about the mean of the differences, or, in other words, the mean of the treatment impacts. But, since the Student's $\underline{t}$ distribution describes the difference of means, it also describes the mean of the differences. The Bayesian interpretation is that we are investigating a stochastic quantity--the mean of the individual impacts of the treatment under discussion. Our knowledge about that quantity is itself probabilistic. That is, our knowledge about the mean individual impact is described by a posterior knowledge distribution, which distribution has its own shape and parameters. Because of the specific premises adopted, that posterior knowledge distribution is a Student's $\underline{t}$ distribution, characterized by three parameters: a degrees of freedom, a mean, and a scale factor.

In this particular situation, although our knowledge about the mean individual impact is still probabilistic, the posterior distribution describing that knowledge is fairly narrow. In other words, the range of plausible values for the mean individual impact is relatively small. The mean individual impact is almost surely greater than -1 point and less than 1 point. It is important to appreciate the limitations as well as the strengths implied by this formulation. The principal strength of the Bayesian presentation is that it gives precise, quantitative, and formal expression to our knowledge about a process parameter of central interest, without falsely implying that that parameter is known exactly. One important limitation of the analysis is that the mean individual impact is not the only process parameter of interest. We shall also wish to express our knowledge about the fluctuation of impacts between individuals. This fluctuation can be indexed by the standard deviation of the individual impacts. We shall discuss this latter notion in a few paragraphs. Before doing so, however, it is important to give some explanations of the features of a Bayesian analysis approach. First, these conclusions can be generalized only to situations in which the analyst judges that all relevant influences and initial conditions are the same as those obtaining here. Thus, the degree of generalizability of the conclusions is one area in which the judgments are still mainly informal. But this, of course, is equally true for conventional analysis.

Because Table 4.3 is the first of many similar tables to be presented throughout this chapter, it is useful to review in some detail its contents.

There are three columns of numbers in the body of Table 4.3, preceded by some heading information which identifies the particular Student's t distribution being tabulated. As already stated, that distribution represents our probabilistic knowledge about the mean of the individual treatment impacts. The Student's t distribution can be shown to be the marginal posterior distribution for the difference of two means, provided each of the means forming the difference comes from a Normal population, and provided that the variance in each of the two Normal populations is the same, through the variance may not be known (cf. Novick and Jackson, 1974; Box and Tiao, 1973).

Although this is the same distribution a sampling-theory analyst might employ in this same situation, our interpretation of it differs from a sampling-theory interpretation in several ways. Some are minor; others are quite important. One of the less important but potentially confusing differences in handling of the Student's t is that we report an unstandardized t distribution. The usual practice among sampling theorists is to rescale the t distribution derived from his data so that it has unit variance, and then to discuss the outcomes with reference to critical values on this standardized Student's t distribution. There is nothing seriously wrong with this approach, but in the present advanced age of computing, there is also little advantage to it. The real interest of the analysis is on the observed variable; if a computer is used to obtain the distribution of the t variable, it is essentially as easy, from a computational viewpoint, to work with the unstandardized t. This is what we have done in the accompanying Tables. As a guide to the user who wishes to relate this to the traditional approach, we report not only the degrees of freedom of the Student's t, and the mean of that same distribution, but also its scale factor. The standardized t distribution has a scale factor of 1.0. A little thought about the common-sense meaning of a scale factor indicates that if a scale factor is small, then a given observed value of t will lie farther out on the tail of the distribution. So, for example, in Table 4.3, the mean of the Student's t is 0.06, and the scale factor is 0.2454. To convert this to a standardized t distribution, we could divide each possible t value by the scale factor. If we did this for the mean of the observed differences (0.06), we would obtain 0.24. This is equal (within the rounding error that has accumulated) to the value of 0.22 given in the original table. Similarly, for the data in Table 4.4, the observed scale factor is .3372, and the observed mean is 1.58. If we rescale by calculating 1.58/.3372, we obtain 4.69, which is equal to the result reported in the original table.

In this unstandardized form, the distribution shown in Table 4.3 describes directly the probabilistic knowledge we have concerning the mean individual impact of the TO treatment. That mean individual impact is not known exactly, but--most clearly in the Bayesian perspective--is known probabilistically. More specifically, we can make the following statement. Because of the particular assumptions we have made about the state of our prior knowledge and about the data-generating process, and taking full account of the data actually observed, our probabilistic knowledge about the mean individual impact is properly represented by the particular Student's t distribution with the following parameters: mean = 0.06; scale factor = 0.2454; and degrees of freedom = 2200. The three main columns of Table 4.3 describe that particular t distribution in some detail.

The left column, labeled "Impact Value", indicates the point at which this particular t distribution is being evaluated. The middle column, labeled "Density", gives the height of the appropriate Student's t density function at that point. Finally, the third column, labeled "Cumulative Probability", shows the cumulative probability that this particular t variable is less than or equal to the point shown in column 1 on the same line.

Several interpretations of this distribution as it describes our knowledge about the mean impact can readily be drawn. First--quite obvious but nevertheless important to state explicitly--the Table indicates that we cannot state with certainty that the mean treatment impact is or is not equal to some value. In fact, we cannot state with certainty even whether or not it is within some particular range of values. Even though our knowledge about the mean impact is based on a large number of cases (2202), it is still probabilistic.

But, although our knowledge does not allow us to draw any deterministic conclusions about the mean individual impact, it does provide the grounds for rather strong probabilistic conclusions. For example, it is clear that the mean impact almost certainly lies within a range of only about one gain point. This is evident from inspection of the entries in column 3 of the Table. Those entries indicate that the probability of the mean impact being less than or equal to -0.5 is .0113, and the probability of the mean impact being greater than +0.5 is .0365 (= 1.0000 - .9635). Just as our intuition would expect, the fact that there are many cases available causes our posterior knowledge distribution to become quite localized. Notice too that the scale factor, 0.2454, provides a single summarizing measure of the width of the distribution under discussion in comparison with other t distributions.

Column 2 of the Table shows the density values of the particular Student's t distribution at each point under consideration. Notice that these densities may be larger than 1.0 at the middle parts of the distribution. These density values can be thought of as standardized conditional likelihood values at the point in question. They can be used to form likelihood ratios for comparing two possible points. More generally, the relative size of the area under different segments of the curve corresponds to the relative probabilities of the mean impact actually lying within the two respective intervals. Again, this is nothing more than the usual relationship between a probability density curve and the likelihoods or the odds of two hypotheses. Notice here, however, that it is being used directly to represent knowledge about the process parameter of main interest.

The next table, Table 4.4, is similar in layout to the one just discussed. It presents corresponding results for the comparison of

TABLE 4.4

READING STANDARD SCORE GAINS, INDIVIDUAL LEVEL ANALYSIS, PARENT-TEACHER PAYMENT, CONTROL VERSUS EXPERIMENTAL

	N	Mean	Variance
Control	775	6.01	39.4384
Experimental	694	7.59	44.0896

Student's t distribution for Impact:

degrees of freedom = 1467 Mean = 1.58 Scale = .3372

(1) Impact Value	(2) Density	(3) Cumulative Probability
-1.0	0.0000	0.0000
-0.5	0.0000	0.0000
0.0	0.0000	0.0000
0.5	0.0071	0.0007
1.0	0.2696	0.0428
1.5	1.1500	0.4062
2.0	0.5445	0.8935
2.5	0.0288	0.9968
3.0	0.0002	1.0000

the EXP and CON students in the PT (Parent-Teacher payment) treatment. This table shows a rather substantial difference in mean gain. The EXP group students, on the average, have gained 1.58 more HBJ score points than the CON group students. This particular Student's t distribution has 1467 degrees of freedom, a mean of 1.58, and a scale factor of .3372. This t distribution thus is located at a higher point, and is somewhat wider, than the one described in Table 4.3. For example, there is a range of approximately 1.5 units required to encompass almost all of the probability. That is, the probability that the mean individual impact lies at or below 1.0 is only 0.0428, and the probability it lies above 2.5 is only 0.0032 (= 1.0000 - .9968). Thus, for this treatment, there seems little reasonable doubt that the EXP group really gained more than the CON group. Again our knowledge is only probabilisitc, but it yields clear evidence in favor of a treatment impact.

To make these results more vivid, the two distributions are presented in graphic form in Figures 4.1 and 4.2, each showing one of the Student's t distributions just discussed.

We turn next to another aspect of the analysis which has been highlighted by the use of a Bayesian perspective. Quite simply, it seems necessary,for several reasons, to allow for the possibility that the impact of either of these treatments may not be constant for all students. Three kinds of reasons can be suggested: (1) the actual implmeentation of the treatment may vary from one teacher or school or grade to another; (2) the student may respond in a strictly

FIGURE 4.1

```
---------------------------------------
READ G-IN, INDIV TO C VS E 15 NOV 76
DUAL -T-   DENSITY ON  2200.0000 D.F., MEAN =    0.0600, SCALE  =    0.2454
X-AXIS IS * IF DENSITY <  0.0325; OTHERWISE IS I
                TICK MARKS ON Y-AXIS EVERY   0.3250

                0                                               1.625
                I----:----I----:----I----:----I----:----I----:----I----:
 1      -1.0000 *
 2      -0.9500 *
 3      -0.9000 *
 4      -0.8500 *
 5      -0.8000 *
 6      -0.7500 *
 7      -0.7000 *
 8      -0.6500 *
 9      -0.6000 I*
 10     -0.5500 I *
 11     -0.5000 I   *
 12     -0.4500 I     *
 13     -0.4000 I        *
 14     -0.3500 I           *
 15     -0.3000 I                *
 16     -0.2500 I                      *
 17     -0.2000 I                            *
 18     -0.1500 I                                  *
 19     -0.1000 I                                       *
 20     -0.0500 I                                            *
 21      0.0000 I                                                *
 22      0.0500 I                                                 *
 23      0.1000 I                                                *
 24      0.1500 I                                              *
 25      0.2000 I                                         *
 26      0.2500 I                                    *
 27      0.3000 I                              *
 28      0.3500 I                        *
 29      0.4000 I                  *
 30      0.4500 I             *
 31      0.5000 I         *
 32      0.5500 I      *
 33      0.6000 I   *
 34      0.6500 I  *
 35      0.7000 I *
 36      0.7500 *
 37      0.8000 *
 38      0.8500 *
 39      0.9000 *
 40      0.9500 *
 41      1.0000 *
---------------------------------------
```

FIGURE 4.2

```
----------------------------------------
READ G-INS INDIV RT C WS E 15 NOV 76
DUAL -T-   DENSITY ON  1467.0000 D.F., MEAN =    1.5800, SCALE  =    0.3372
X-AXIS IS * IF DENSITY >  0.0237; OTHERWISE IS I
                    TICK MARKS ON Y-AXIS EVERY   0.2370
                                                                      1.185
                    0
                    I----:----I----:----I----:----I----:----I----:----I----:
  1          0.2000 *
  2          0.2500 *
  3          0.3000 *
  4          0.3500 *
  5          0.4000 *
  6          0.4500 *
  7          0.5000 *
  8          0.5500 *
  9          0.6000 *
 10          0.6500 I*
 11          0.7000 I *
 12          0.7500 I *
 13          0.8000 I  *
 14          0.8500 I    *
 15          0.9000 I      *
 16          0.9500 I        *
 17          1.0000 I          *
 18          1.0500 I              *
 19          1.1000 I                 *
 20          1.1500 I                     *
 21          1.2000 I                         *
 22          1.2500 I                              *
 23          1.3000 I                                  *
 24          1.3500 I                                       *
 25          1.4000 I                                          *
 26          1.4500 I                                             *
 27          1.5000 I                                                *
 28          1.5500 I                                                 *
 29          1.6000 I                                                 *
 30          1.6500 I                                                *
 31          1.7000 I                                              *
 32          1.7500 I                                           *
 33          1.8000 I                                       *
 34          1.8500 I                                   *
 35          1.9000 I                               *
 36          1.9500 I                          *
 37          2.0000 I                      *
 38          2.0500 I                  *
 39          2.1000 I              *
 40          2.1500 I           *
 41          2.2000 I        *
 42          2.2500 I      *
 43          2.3000 I    *
 44          2.3500 I   *
 45          2.4000 I  *
 46          2.4500 I *
 47          2.5000 I*
 48          2.5500 *
 49          2.6000 *
 50          2.6500 *
 51          2.7000 *
 52          2.7500 *
 53          2.8000 *
----------------------------------------
```

stochastic way to even an identical treatment; and (3) there may be an "aptitude-treatment" interaction. Yet, our analysis thus far has effectively discussed only the mean individual impact. We have ignored any discussion of the fluctuation in the impact.

Some careful reflection indicates that this question--the size of the fluctuation in treatment impact--cannot be addressed very well with this particular experimental design. Yet, the fragmentary pieces of evidence we can discern in the data suggest that it does occur, and is probably sizeable. For example, within a single site, there is considerable fluctuation between grade levels in the average size of the measured treatment impact. Also, in all cases where a treatment impact does occur as indicated by the difference in mean gains, there is a corresponding difference between the standard deviations of the two groups. This suggests the likely presence of a de-homogenizing process of some sort, but the summary data available to us do not allow any very adequate pursuit of the topic. Thus, we simply note it in passing as an incidental payoff of a Bayesian approach--that of sensitizing us to the real importance of some of the assumptions we make.

Primarily for illustrative purposes, we present next some Tables showing the posterior knowledge distribution for knowledge about the process standard deviation. These Tables correspond to the similar Tables given already showing distributions of the mean impact. Table 4.5 shows, for a particular set of assumptions, the posterior SCL of the standard deviation of the gains in score for

TABLE 4.5

Inverse Chi Posterior Knowledge Distribution for the Standard Deviation of Reading Gains, among all students in the TO study (EXP and CON combined)

Mode = 5.7514 Scale Factor = 269.8269 DF = 2200

Std. Deviation Value	Density	Cumulative Probability
5.50	0.0495	.0012
5.60	0.9334	.0351
5.70	3.8479	.2667
5.80	3.9403	.7024
5.90	1.1256	.9508
6.00	0.0997	.9970
6.10	0.0030	.9999

TABLE 4.6

Inverse Chi Posterior Knowledge Distribution for the Standard Deviation of Reading Gains, among all students in the PT study (EXP and CON combined)

Mode = 6.4500 Scale Factor = 247.1282 DF = 1467

Std. Deviation Value	Density	Cumulative Probability
6.00	0.0011	.0000
6.10	0.0292	.0009
6.20	0.3176	.0139
6.30	1.4669	.0942
6.40	3.0631	.3251
6.50	3.0699	.6505
6.60	1.5598	.8860
6.70	0.4226	.9774
6.80	0.0639	.9973

students in the TO treatment condition. Table 4.6 shows the same distribution for the gains in score for students in the PT treatment condition.

The contents of these tables indicate the particular, quantified form of the probabilistic knowledge we have, as a result of some cautious initial beliefs, and the analysis of the data from this study, as to the possible values of the standard deviations of gains scores for students. Under the assumptions to be described, these posterior knowledge distributions follow the inverse chi distribution (cf. Novick and Jackson, 1974; Box and Tiao, 1973). As Table 4.5 shows, it is almost certain that the standard deviation of Reading gains for students in the Teacher-Only study lies between 5.5 and 6.1. Similarly, it is also almost certain that the standard deviation of Reading gains for students in the Parent-Teacher study lies between 6.1 and 6.8. Neither of these standard deviations is known exactly. Yet, the sample evidence in both studies is based upon a large number of individuals (2202 students in the TO study, and 1469 students in the PT study). This sizeable quantity of evidence allows us to specify our knowledge about these sigmas in the form of a narrow probability distribution (SCL).

The preceding analyses using the Bayesian perspective have concerned only the Reading scores. We turn now to the presentation of a corresponding set of reanalysis results for the Mathematics scores. Tables 4.7 and 4.8 present the Student's t posterior knowledge distributions for the treatment impact on Math gains, first for the TO

TABLE 4.7

MATH STANDARD SCORE GAINS, INDIVIDUAL LEVEL ANALYSIS, TEACHER-ONLY PAYMENT, CONTROL VERSUS EXPERIMENTAL

	N	Mean	Variance
Control	1037	8.79	58.0644
Experimental	1144	9.38	60.9961

Student's t distribution for Impact:

degrees of freedom = 2179 Mean = 0.59 Scale Factor = 0.3310

(1) Impact Value	(2) Density	(3) Cumulative Probability
-1.00	0.0000	0.0000
-0.50	0.0054	0.0005
0.00	0.2462	0.0374
0.50	1.1613	0.3928
1.00	0.5596	0.8922
1.50	0.0277	0.9970
2.00	0.0001	1.0000
2.50	0.0000	1.0000
3.00	0.0000	1.0000

TABLE 4.8

MATH STANDARD SCORE GAINS, INDIVIDUAL LEVEL ANALYSIS, PARENT-TEACHER PAYMENT, CONTROL VERSUS EXPERIMENTAL

	N	Mean	Variance
Control	775	7.51	63.0436
Experimental	691	11.29	80.2816

Student's t distribution for Impact:

degrees of freedom = 1464 Mean = 3.78 Scale Factor = 0.4414

(1) Impact Value	(2) Density	(3) Cumulative Probability
-1.00	0.0000	0.0000
-0.50	0.0000	0.0000
0.00	0.0000	0.0000
0.50	0.0000	0.0000
1.00	0.0000	0.0000
1.50	0.0000	0.0000
2.00	0.0003	0.0000
2.50	0.0136	0.0019
3.00	0.1897	0.0387
3.50	0.7389	0.2629
4.00	0.7981	0.6909
4.50	0.2390	0.9485
5.00	0.0200	0.9972
5.50	0.0005	1.0000

(Teacher-Only) treatment, and then for the PT (Parent-Teacher) treatment. Inspection of Table 4.7 indicates that knowledge about the mean individual impact on math scores is (as with reading scores) probabilistic. The impact is almost certainly positive (i.e., its probability of being less than or equal to zero is .0374), but also almost certainly less than 1.50 (HBJ) score points (its probability of being greater than 1.5 score points is .0030). Thus, on math, the TO treatment seems overall to have some small positive mean impact. Turning next to Table 4.8, we find a clearly positive and sizeable impact. That is, here too our knowledge about the impact of the PT treatment is probabilistic, but the plausible values of the mean individual impact lie between 2.50 and 5.00, all of which are sizeable and positive.

Tables 4.9 and 4.10 present the posterior standardized conditional likelihood curves for our knowledge about the standard deviation of the gains scores in Math, among the students in the TO treatment, and then among the students in the PT treatment. These tables reflect the existence of reasonably clear though probabilistic knowledge about the value of the standard deviations of the gains. For students in the TO (Teacher-Only) treatment, Table 4.9 indicates that the standard deviation of the gains almost certainly lies between 7.4 and 8.0. Similarly, for the students in the PT (Parent-Teacher) treatment, Table 4.10 indicates that the standard deviation of the Math gains almost certainly lies between 8.1 and 8.9.

TABLE 4.9

Inverse Chi Posterior Knowledge Distribution for the Standard Deviation of Math Gains, among all Students in the TO Study (EXP and CON combined)

Mode = 7.7180 Scale Factor = 360.3555 DF = 2179

Std. Deviation Value	Density	Cumulative Probability
7.20	0.0001	0.0000
7.30	0.0031	0.0001
7.40	0.0645	0.0022
7.50	0.5507	0.0265
7.60	2.0232	0.1474
7.70	3.3715	0.4289
7.80	2.6784	0.7487
7.90	1.0626	0.9335
8.00	0.2198	0.9897
8.10	0.0247	0.9991
8.20	0.0016	1.0000

TABLE 4-10

Inverse Chi Posterior Knowledge Distribution for the Standard Deviation of Math Gains, among all Students in the PT Study (EXP and CON combined)

Mode = 8.4335 Scale Factor = 322.7942 DF = 1464

Std. Deviation Value	Density	Cumulative Probability
7.90	0.0037	0.0001
8.00	0.0374	0.0017
8.10	0.2208	0.0124
8.20	0.7889	0.0588
8.30	1.7554	0.1844
8.40	2.5002	0.4027
8.50	2.3388	0.6530
8.60	1.4722	0.8463
8.70	0.6379	0.9491
8.80	0.1943	0.9874
8.90	0.0425	0.9977
9.00	0.0068	0.9997
9.10	0.0008	1.0000

To conclude this section of the reanalysis, we present next a series of Tables indicating the posterior knowledge we have about the overall impact of the incentives strategy. That is, we combine the results from the TO study and those from the PT study, and interpret these combined results as reflecting the mean impact, and the amount of the standard deviation of the gains, for the general strategy of making incentive payments to the adults involved in the operations of a school. Notice that one use of the estimated overall standard deviation of the gains is to provide a frame of reference by which to interpret the magnitude of the mean impacts. In other words, we might note that the mean posterior knowledge distribution for the impact of the PT treatment on Math scores is 3.78. At the same time, our study of the posterior distribution for the standard deviation of those gains indicates that it is likely to lie in the neighborhood of 8.4. This in turn suggests that the mean individual impact of the PT treatment on the Math outcomes is approximately one-half a standard deviation. This is a sizeable impact, though not enormous. Considering the simplicity of the treatment, and the relatively chaotic way it was operated in this field trial, these results are strongly encouraging as to the viability of such an approach.

As a general overall result, Tables 4.11 and 4.12 present our posterior knowledge about the mean individual impact, and about the standard deviation of the gains, for the Reading scores, and for all students in the TO and the PT treatments combined. These results

TABLE 4.11

READING STANDARD SCORE GAIN, INDIVIDUAL ANALYSIS, ALL EXPERIMENTAL VERSUS ALL CONTROL

	N	Mean	Variance
Control	1830	6.25	35.4025
Experimental	1841	6.90	37.9456

Student's t distribution for Impact:

degrees of freedom = 3669 Mean = 0.65 Scale Factor = 0.1999

(1) Impact Value	(2) Density	(3) Cumulative Probability
-1.00	0.0000	0.0000
-0.50	0.0000	0.0000
0.00	0.0102	0.0005
0.50	1.5058	0.2265
1.00	0.4311	0.9600
1.50	0.0002	1.0000
2.00	0.0000	1.0000
2.50	0.0000	1.0000
3.00	0.0000	1.0000

TABLE 4.12

Inverse Chi Posterior Knowledge Distribution for the Standard Deviation of Reading Gains, among all EXP and all CON Students Combined

Mode = 6.0550 Scale Factor = 366.8148 DF = 3669

Std. Deviation Values	Density	Cumulative Probability
5.85	0.0656	0.0013
5.90	0.4581	0.0119
5.95	1.8117	0.0634
6.00	4.1499	0.2108
6.05	5.6292	0.4641
6.10	4.6192	0.7302
6.15	2.3399	0.9045
6.20	0.7461	0.9768
6.25	0.1526	0.9962
6.30	0.0204	0.9996
6.35	0.0018	1.0000

indicate overall a small positive impact of the treatment, almost certainly between 0.0 and 1.0 HBJ score points. This impact is indeed not large, but at the same time it is clearly in the hoped-for direction. The results of Table 4.12 indicate that the overall standard deviation of gains lies in the region near 6.05.

Tables 4.13 and 4.14 present corresponding results of the overall incentives strategy on Math outcomes. Here, as Table 4.13 indicates, the overall mean individual impact of the treatments is certainly positive, and almost certainly lies between approximately 1.0 and 3.0 score points in magnitude. Table 4.14 indicates that the overall standard deviation of the gains is in the neighborhood of 7.9 to 8.2. These results seem quite clearly to indicate that the incentive strategy deserves further investigation on mathematics outcomes.

4.5 Reanalysis at the Level of Grades

Although achievement measurements were made on individual students, the unit of payment was the classroom, and thus this is, for many purposes, the more meaningful unit of analysis. If we consider again the impact of the PT (Parent-Teacher) payment program, viewing classrooms as the logically correct unit, we find technical obstacles interfere with the feasibility of a reanalysis at that level. It is possible, however, to use grade level units as an approximation to the classroom units. The approximation will tend to dampen out some sources of fluctuation, but is still worth exploration. In Table 4.15 we present a more summarized version of the Student's t analysis, where grades are regarded as the units. This choice of unit implies

TABLE 4.13

MATH STANDARD SCORE GAINS, INDIVIDUAL LEVEL ANALYSIS, ALL EXPERIMENTAL VERSUS ALL CONTROL

	N	Mean	Variance
Control	1812	8.24	60.5284
Experimental	1835	10.10	69.0561

Student's _t_ distribution for Impact:

degrees of freedom = 3645 Mean = 1.86 Scale Factor = 0.2666

(1) Impact Value	(2) Density	(3) Cumulative Probability
-1.00	0.0000	0.0000
-0.50	0.0000	0.0000
0.00	0.0000	0.0000
0.50	0.0000	0.0000
1.00	0.0083	0.0006
1.50	0.6013	0.0885
2.00	1.3034	0.7003
2.50	0.0841	0.9918
3.00	0.0002	1.0000

TABLE 4.14

Inverse Chi Posterior Knowledge Distribution for the Standard Deviation of Math Gains, among all EXP and all CON Students Combined

Mode = 8.0488   Scale Factor = 486.0021   DF = 3645

Std. Deviation Values	Density	Cumulative Probability
7.60	0.0000	0.0000
7.70	0.0027	0.0001
7.80	0.1078	0.0032
7.90	1.1706	0.0523
8.00	3.6962	0.2947
8.10	3.6562	0.6988
8.20	1.2151	0.9406
8.30	0.1448	0.9950
8.40	0.0066	0.9998
8.50	0.0001	1.0000

TABLE 4.15

READING STANDARD SCORE IMPACT, GRADE LEVEL ANALYSIS, TEACHER-ONLY PAYMENT, EXPERIMENTAL VERSUS CONTROL

Student's t distribution for Impact:

degrees of freedom = 11 Mean = -.1315 Scale Factor = .5231

Impact Value	Density	Cumulative Probability
-1.00	.1949	.0625
-0.50	.5721	.2479
0.00	.7203	.5969
0.50	.3534	.8737
1.00	.0889	.9733
1.50	.0167	.9951
2.00	.0030	.9991
2.50	.0006	.9998
3.00	.0001	1.0000

that the Student's _t for differences_ is based on 24 cases. There are 12 grades in the EXP condition, and 12 in the CON condition. The grades are individually paired by site and grade level, and so the analysis comes to be based on 12 cases. This leads to a Student's _t_ distribution for the treatment impact based on 11 degrees of freedom. The surprising feature here is that with this quite reasonable interpretation, we abruptly find ourselves in a "small-sample" analysis. The results given in Table 4.15 are calculated using an ignorance or indifference prior, and they also use the observed spread among the grade level means as the basis for deducing the scale factor of the Student's _t_.

Table 4.15, based on 11 degrees of freedom, indicates a fact not clear from the individual level analyses previously presented. There _is_ a sizeable degree of between-grade fluctuation in the estimated mean impact, so that the posterior knowledge distribution about the mean impact is made more diffuse. In general, we may be reasonably sure that the impact is not substantial in either direction. That is the probability of a negative impact tending to reduce the EXP gain more than 1 (HBJ) score unit is approximately .063, but correspondingly the probability of a treatment impact tending to raise the EXP gain more than 1 (HBJ) score unit also is small, approximately .027. Between those limits, however, the chances of a negative impact are higher than for a positive impact. In short, the data suggest that the TO treatment has no consistent effects on Reading gains, but that negative changes occur slightly more often than positive ones.

If we examine the corresponding analysis for Math gains, presented in Table 4.16, we find a slightly more positive, but essentially similar, outcome. That table indicates, for example, that the chances of a sizeable effect in either direction (greater than 1 score unit) is small (.024 for a negative difference, .072 for a positive difference). Here, the odds favor a positive outcome, .624/.376, but again the range of plausible values is quite a bit larger than was the case with the individual analysis.

This perhaps paradoxical-appearing result arises not simply because we have fewer cases in the grade-level analysis, but also, and importantly, because a larger-than-chance fraction of the fluctuation in individual impacts does lie between grades. This is, of course, simply an indication that grade level (and by implication classroom) is a substantively real category. Treatment inpacts are relatively homogeneous within each grade (though still diverse), and do show the effects of real differences between grade.

There are further implications of these results as well, but before discussing them we first present, in Tables 4.17 and 4.18, the corresponding results for the PT treatment, for Reading and Math respectively.

The results for the PT treatment on Reading indicate that the impact is most probably positive (.07 chance that it is less than zero), and that it may be sizeable (greater than 2 (HBJ) score units). The probability of an impact greater than 2 score units is about .11. For the Math scores, the results are strongly positive. The impact

TABLE 4.16

MATH STANDARD SCORE IMPACT, GRADE LEVEL ANALYSIS, TEACHER-ONLY PAYMENT, EXPERIMENTAL VERSUS CONTROL

Student's t distribution for Impact:

degrees of freedom = 11 Mean = 0.1704 Scale Factor = .5272

Impact Value	Density	Cumulative Probability
-1.00	.0802	.0242
-0.50	.3249	.1149
0.00	.6989	.3763
0.50	.5999	.7277
1.00	.2188	.9280
1.50	.0479	.9858
2.00	.0088	.9974
2.50	.0016	.9995
3.00	.0003	.9999

TABLE 4.17

READING STANDARD SCORE GAINS, GRADE LEVEL ANALYSIS PARENT-TEACHER PAYMENT, EXPERIMENTAL VERSUS CONTROL

Student's _t_ distribution for Impact:

degrees of freedom = 11 Mean = 1.1056 Scale Factor = .7007

Impact Value	Density	Cumulative Probability
-1.00	.0153	.0060
-0.50	.0536	.0213
0.00	.1637	.0715
0.50	.3753	.2030
1.00	.5497	.4415
1.50	.4694	.7076
2.00	.2430	.8860
2.50	.0880	.9640
3.00	.0262	.9897

TABLE 4.18

MATH STANDARD SCORE GAINS, GRADE LEVEL ANALYSIS PARENT-TEACHER PAYMENT, EXPERIMENTAL VERSUS CONTROL

Student's t distribution for Impact:

degrees of freedom = 11 Mean = 2.9665 Scale Factor = .6413

Impact Value	Density	Cumulative Probability
-1.00	.0001	.0000
-0.50	.0003	.0001
0.00	.0009	.0004
0.50	.0037	.0014
1.00	.0149	.0054
1.50	.0590	.0215
2.00	.1972	.0800
2.50	.4587	.2411
3.00	.6072	.5204
3.50	.4217	.7884
4.00	.1705	.9323
4.50	.0494	.9821
5.00	.0124	.9955
5.50	.0030	.9989
6.00	.0008	.9997
6.50	.0002	.9999
7.00	.0001	1.0000

is almost certainly greater than plus 1 unit and has a .48 chance of being greater than 3.0 (HBJ) score units. The range of plausible values of the impact here again is larger than it was for the individual level analysis.

These results suggest that the Incentives Study, though relatively major in scope by comparison with other efforts aimed at similar purposes, was in fact a rather small study in terms of the number of cases involved, if we choose to regard grades as the units. They further reveal that the mean impact of each treatment, on each subject area, is not consistent. This latter conclusion is particularly important if the study results are considered for use in policy. The implication is, of course, present in the data rather than the method of analysis. It is merely highlighted by the Bayesian emphasis on knowledge of parameters as distributional. In this case, our SCL distributions are rather wide-spread.

4.6 Analysis at the Level of Schools

Just as the preceding section makes a reasonable claim that classrooms are one logical choice as the unit of analysis (and then uses grades as a proxy for classrooms), so an argument can be made that, from the point of view of a policy planner in a State or Federal agency, the relevant unit is the entire school. The implementation of any programs such as the Incentives payments is feasible only at the school level, and so one would wish to assess the impact of such programs, considering entire schools as the units. The attempt to use such reasoning, however, immediately makes clear how costly such research could be, and how limited our present knowledge is. To use schools as the analysis unit, we find there are only 4 units in the TO experiment, arranged in 2 pairs. This in turn implies that the degrees of freedom available for the Student's _t_ distribution at the school level has shrunken to 1 degree of freedom. In short, this study is the logical minimum one that could be done for such purposes.

Not surprisingly, the small degrees of freedom tends to diffuse further our posterior knowledge about the mean impact. As Tables 4.19 and 4.20 show, for Reading and Math respectively, the impact of the TO treatment is known only in outline. For Reading there is no clear evidence that the TO treatment has any systematic effect. For Math, the situation is similar, but slightly more favorable. That is, the SCL distributions of possible impact values both are spread rather widely, but the one for Math has a slightly higher mean. Clearly, the limitations on our inference here are a result mainly of the combination of no major impact coupled with few cases.

The situation for the impact of the PT treatment is different. This treatment has a fairly widespread impact on Reading, and one of moderate size on the average. On Math, however, the impact is clear and large. It is almost certainly greater than 2 score units, and very likely greater than 3 score units. These results are consistent numerically with those from the original analysis, but are stronger in the interpretation that can be given. This is especially true for the instances where the results would be, by sampling-theory analysis, not significant, or just barely significant. The Bayesian approach

TABLE 4.19

READING STANDARD SCORE GAINS, SCHOOL LEVEL ANALYSIS TEACHER-ONLY PAYMENT, EXPERIMENTAL VERSUS CONTROL

Student's _t_ distribution for impact:

degrees of freedom = 1 Mean = -0.1100 Scale Factor = .5700

Impact Value	Density	Cumulative Probability
-3.00	.0209	.0620
-2.50	.0301	.0745
-2.00	.0466	.0932
-1.50	.0804	.1239
-1.00	.1624	.1813
-0.50	.3804	.3090
0.00	.5384	.5607
0.50	.2603	.7608
1.00	.1165	.8490
1.50	.0622	.8917
2.00	.0380	.9160
2.50	.0254	.9316
3.00	.0181	.9423
3.50	.0136	.9502
4.00	.0105	.9561
4.50	.0084	.9608
5.00	.0069	.9646
5.50	.0057	.9678
6.00	.0048	.9704

TABLE 4.20

MATH STANDARD SCORE GAINS, SCHOOL LEVEL ANALYSIS, TEACHER-ONLY PAYMENT, EXPERIMENTAL VERSUS CONTROL

Student's t distribution for Impact:

degrees of freedom = 1 Mean = .2100 Scale Factor = 1.3200

Impact Value	Density	Cumulative Probability
-5.00	.0145	.0790
-4.50	.0176	.0870
-4.00	.0216	.0967
[illegible].50	.0271	.1088
-3.00	.0349	.1242
-2.50	.0462	.1443
-2.00	.0634	.1714
-1.50	.0900	.2092
-1.00	.1310	.2638
-0.50	.1870	.3429
0.00	.2352	.4498
0.50	.2300	.5689
1.00	.1775	.6717
1.50	.1233	.7463
2.00	.0849	.7978
2.50	.0601	.8336
3.00	.0441	.8593
3.50	.0334	.8786
4.00	.0261	.8933
4.50	.0209	.9050
5.00	.0170	.9144
5.50	.0141	.9222
6.00	.0119	.9287

indicates the plausibility of several interpretations simultaneously, not just the plausibility of one rather irrelevant hypothesis (the hypothesis that there is no impact).

From these reanalyses, using the Bayesian viewpoint, we see that the data imply that a range of outcomes is to be expected. This range means that there is no way we can meaningfully speak about "the" impact of a treatment. Rather, we must describe our best judgment as to the distribution of impacts that may be expected. Beyond the idea of impact as distributional, the Bayesian approach also encouraged us to pursue the analysis at several different levels of aggregation. Using each of these is appropriate for slightly different research purposes. However, when units such as schools are chosen (a choice which corresponds to the Federal policy-former's needs) there is a serious difficulty in obtaining data which is simultaneously high in quality and quantity. The Incentives project had data of rather good quality, but on only a few schools.

This reanalysis of the Incentives data has tended to show that the original results did point in the proper directions, but were not clearly enough delineated. It indicates also that the impacts of each treatment are best conceived as being variable, with components of variability arising from differences between individuals, from differences between classrooms and grades, and from differences between schools. This latter notion was implicit in the sets of outcome mean scores presented in the original analysis, but was not made coherent nor clearly represented.

4.7 Sensitivity to Prior Beliefs

We turn now to the reanalysis involving the second major feature of the Bayesian perspective, the formalized incorporation and synthesis of prior beliefs.

The central justifying idea here is simply that persons do not take research results and accept them as if those results are the only source of guidance as to the characteristics of the situation. Instead, they blend such results together with their previous beliefs and develop a new composite judgment. This process in practice occurs privately and informally, perhaps even without conscious awareness by the person. Yet it is reasonable and prudent as a general strategy.

The thrust of the Bayesian approach in this connection is that it provides an explicit, formal, and normatively correct algorithm for this blending of previously formed beliefs with new evidence.

Depending upon the various aspects of these two inputs, various blends can be expected. To assess in a general way the sensitivity of the final judgment to differences in the content of prior beliefs, while holding constant the data actually observed, we carried out a series of combinations of different priors with the Incentives project data.

These analyses were confined to the grade level and school level. This restriction was adopted because the Bayesian formulas in general weight prior beliefs and sample evidence in proportion to the number of cases each reflects. At the individual level, the number of sample cases is high enough that the study evidence would tend to dominate any reasonable assignment of case size to a teacher's prior beliefs.

Three distinct characteristics of the prior beliefs are important, and each of these three was varied systematically through several values. These three characteristics are (1) the number of cases equivalently represented by the judged values; (2) the judged mean impact; and (3) the judged standard deviation of those impacts across units. These same three aspects are important in the data as well. The data-based analyses presented previously included a distributional judgment about values of the mean impact, and another distributional judgment about the standard deviation of the impacts. Each of those distributions was in part determined by the number of cases in the sample. An exactly analogous process applies to the prior beliefs.

For the analysis of sensitivity to prior beliefs at the grade level unit of analysis, the premise was adopted that a person choosing this unit most probably would be a school principal or curriculum specialist, and would be likely to feel his relevant prior experience involved between 6 and 24 grades. Since this analysis is being used here mainly as a substitute for a classroom level of analysis, this suggests between 15-20 and 60-70 classrooms as the experience base.

Workers on the original study indicate in the report and in personal communication that educators at the local sites varied quite widely in their expectations as to the impact the project would have. There were a number who found the incentives idea highly appealing, and who believed that the treatments would lead to a sizeable increase in achievement gain. Others were less optimistic, and felt that a modest shift was the likely outcome. Still others, fairly numerous, felt the incentives idea was misguided and dangerous. They expected that it would actually lead to sizeable negative changes in test score gain.

The third aspect of the prior beliefs is the least intuitive. It refers simply to the judgment as to how much variability there is in the process generating differences of means. It is distinct from the number of cases idea mentioned above. The distinctions here can be communicated best by phrasing them in terms of a Normal distribution. The process operating on the grades is formally modeled by describing the model process as one generating Normally distributed data. The objective is to obtain estimates of the two parameters of that Normal distribution, its mean and its standard deviation. The knowledge of these two estimates will be distributional, and the two knowledge (judgment) distributions will tend to be narrower if based on more cases.

Thus, we specified, in addition to the number of cases and the estimated mean, an estimated standard deviation. The three quantities then were permuted through a series of combinations. For the analyses of outcomes at the grade level, there were three levels of case base, six levels of mean, and three levels of standard deviation, yielding 54 combinations. Each one of these 54 combinations represents one reasonable formalization of prior beliefs about the impacts of the treatment.

These 54 patterns of prior belief were combined with the sample data for the TO treatment, the PT treatment, and the combined treatment, and separately for Reading and for Math. The results are shown in Tables 4-21 through 4-26 which appear below.

Consider first Table 4.21, which shows the sensitivity of the results for the teacher-only treatment to the different configurations of prior judgment. The sample data for the comparison was based on 12 cases, with the mean impact being -0.13, and the sample value of the process standard deviation (unbiased estimate) equal to 1.8122. There are several noteworthy aspects in Table 4.21.

First, consider the person whose prior judgment was that the treatment would be highly effective. This is represented here by setting the prior mean equal to 3.0. Such a person might have differing beliefs about the general variability of outcomes--as reflected in choices of process σ at 0.5, or 1.0, or 1.5. Also, he might have varying degrees of confidence in his prior judgment relative to the sample data. We incorporated this by choosing three alternate equivalent prior cases. These were chosen as twice as influential as the sample, the same as the sample, and half as influential as the sample (N = 24, 12, and 6, respectively).

One feature of this set of computations is that generally, all of the prior judgments about the process variability (standard deviation) were smaller than the estimated standard deviation based on the data. Thus, one's first reaction might be to assume that the composite standard deviation derived from the blend of the prior judgment

TABLE 4.21

READING OUTCOMES SENSITIVITY ANALYSIS
GRADES AS UNITS, TEACHER-ONLY TREATMENT

	prior		prior N = 24 posterior		prior N = 12 posterior		prior N = 6 posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	1.9562	1.8812	1.4342	2.1074	.9123	2.1880
2	3	1.0	1.9562	2.0115	1.4342	2.1946	.9123	2.2409
3	3	1.5	1.9562	2.2117	1.4342	2.3327	.9123	2.3265
4	2	.5	1.2895	1.5167	.9342	1.7338	.5790	1.8632
5	2	1.0	1.2895	1.6760	.9342	1.8388	.5790	1.9251
6	2	1.5	1.2895	1.9114	.9342	2.0015	.5790	2.0240
7	1	.5	.6228	1.2381	.4342	1.4547	.2457	1.6297
8	1	1.0	.6228	1.4284	.4342	1.5784	.2457	1.7001
9	1	1.5	.6228	1.6987	.4342	1.7653	.2457	1.8114
10	0	.5	-.0438	1.1116	-.0657	1.3311	-.0877	1.5298
11	0	1.0	-.0438	1.3202	-.0657	1.4552	-.0877	1.6045
12	0	1.5	-.0438	1.6090	-.0657	1.6648	-.0877	1.7220
13	-1	.5	-.7105	1.1871	-.5657	1.4046	-.4210	1.5888
14	-1	1.0	-.7105	1.3843	-.5657	1.5322	-.4210	1.6610
15	-1	1.5	-.7105	1.6619	-.5657	1.7242	-.4210	1.7747
16	-2	.5	-1.3772	1.4328	-1.0657	1.6489	-.7543	1.7913
17	-2	1.0	-1.3772	1.6001	-1.0657	1.7591	-.7543	1.8555
18	-2	1.5	-1.3772	1.8456	-1.0657	1.9285	-.7543	1.9580

TABLE 4.22

READING OUTCOMES SENSITIVITY ANALYSIS
GRADES AS UNITS, PARENT TEACHER TREATMENT

	prior		prior N = 24 posterior		prior N = 12 posterior		prior N = 6 posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.3685	1.7088	2.0528	2.0124	1.7370	2.2419
2	3	1.0	2.3685	1.8513	2.0528	2.1035	1.7370	2.2937
3	3	1.5	2.3685	2.0671	2.0528	2.2472	1.7370	2.3773
4	2	.5	1.7019	1.5045	1.5528	1.8137	1.4037	2.0807
5	2	1.0	1.7019	1.6647	1.5528	1.9143	1.4037	2.1363
6	2	1.5	1.7019	1.9018	1.5528	2.0711	1.4037	2.2259
7	1	.5	1.0352	1.4415	1.0528	1.7533	1.0704	2.0327
8	1	1.0	1.0352	1.6079	1.0528	1.8572	1.0704	2.0896
9	1	1.5	1.0352	1.8524	1.0528	2.0185	1.0704	2.1810
10	0	.5	0.3685	1.5372	0.5528	1.8452	0.7370	2.1059
11	0	1.0	0.3685	1.6943	0.5528	1.9441	0.7370	2.1608
12	0	1.5	0.3685	1.9277	0.5528	2.0987	0.7370	2.2494
13	-1	.5	-.2981	1.7660	0.0528	2.0689	0.4037	2.2885
14	-1	1.0	-.2981	1.9042	0.0528	2.1576	0.4037	2.3392
15	-1	1.5	-.2981	2.1146	0.0528	2.2979	0.4037	2.4213
16	-2	.5	-.9648	2.0844	-.4472	2.3878	0.0704	2.5575
17	-2	1.0	-.9648	2.2028	-.4472	2.4651	0.0704	2.6028
18	-2	1.5	-.9648	2.3870	-.4472	2.5887	0.0704	2.6768

TABLE 4.23

READING OUTCOMES SENSITIVITY ANALYSIS GRADES AS UNITS, COMBINED TREATMENTS

			prior N = 24		prior N = 12		prior N = 6	
	prior		posterior		posterior		posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	1.7435	2.0411	1.3247	2.1921	0.9896	2.2495
2	3	1.0	1.7435	2.1311	1.3247	2.2468	0.9896	2.2790
3	3	1.5	1.7435	2.2730	1.3247	2.3351	0.9896	2.3275
4	2	.5	1.2435	1.7653	0.9914	1.9643	0.7896	2.0905
5	2	1.0	1.2435	1.8685	0.9914	2.0251	0.7896	2.1222
6	2	1.5	1.2435	2.0288	0.9914	2.1226	0.7896	2.1742
7	1	.5	0.7435	1.6086	0.6580	1.8390	0.5896	2.0057
8	1	1.0	0.7435	1.7212	0.6580	1.9038	0.5896	2.0388
9	1	1.5	0.7435	1.8941	0.6580	2.0072	0.5896	2.0928
10	0	.5	0.2435	1.6065	0.3247	1.8373	0.3896	2.0046
11	0	1.0	0.2435	1.7193	0.3247	1.9022	0.3896	2.0377
12	0	1.5	0.2435	1.8923	0.3247	2.0057	0.3896	2.0918
13	-1	.5	-.2565	1.7595	-.0086	1.9596	0.1896	2.0873
14	-1	1.0	-.2565	1.8630	-.0086	2.0206	0.1896	2.1191
15	-1	1.5	-.2565	2.0238	-.0086	2.1183	0.1896	2.1711
16	-2	.5	-.7565	2.0328	-.3420	2.1852	-.0104	2.2445
17	-2	1.0	-.7565	2.1231	-.3420	2.2400	-.0104	2.2742
18	-2	1.5	-.7565	2.2655	-.3420	2.3286	-.0104	2.3227

TABLE 4.24

MATH OUTCOMES, SENSITIVITY ANALYSIS
GRADES AS UNITS, TEACHER-ONLY TREATMENT

	prior		prior N = 24		prior N = 12		prior N = 6	
			posterior		posterior		posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.0563	1.7698	1.5852	1.9941	1.1136	2.0911
2	3	1.0	2.0563	1.9078	1.5852	2.0861	1.1136	2.1465
3	3	1.5	2.0563	2.1178	1.5852	2.2308	1.1136	2.2357
4	2	.5	1.3901	1.4268	1.0852	1.6450	0.7803	1.7912
5	2	1.0	1.3901	1.5947	1.0852	1.7552	0.7803	1.8554
6	2	1.5	1.3901	1.8409	1.0852	1.9250	0.7803	1.9579
7	1	.5	0.7235	1.1875	0.5852	1.4073	0.4469	1.5948
8	1	1.0	0.7235	1.3847	0.5852	1.5348	0.4469	1.6667
9	1	1.5	0.7235	1.6623	0.5852	1.7264	0.4469	1.7800
10	0	.5	0.0568	1.1203	0.0852	1.3419	0.1136	1.5422
11	0	1.0	0.0568	1.3276	0.0852	1.4750	0.1136	1.6165
12	0	1.5	0.0568	1.6149	0.0852	1.6736	0.1136	1.7331
13	-1	.5	-.6099	1.2532	-.4148	1.4719	-.2197	1.6474
14	-1	1.0	-.6099	1.4415	-.4148	1.5941	-.2197	1.7171
15	-1	1.5	-.6099	1.7098	-.4148	1.7794	-.2197	1.8273
16	-2	.5	-1.2765	1.5352	-.9148	1.7543	-.5531	1.8838
17	-2	1.0	-1.2765	1.6924	-.9148	1.8581	-.5531	1.9451
18	-2	1.5	-1.2765	1.9260	-.9148	2.0193	-.5531	2.0431

TABLE 4.25

MATH OUTCOMES, SENSITIVITY ANALYSIS
GRADES AS UNITS, PARENT-TEACHER TREATMENT

			prior N = 24		prior N = 12		prior N = 6	
	prior		posterior		posterior		posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.9888	1.3290	2.9832	1.6103	2.9777	1.8632
2	3	1.0	2.9888	1.5078	2.9832	1.7229	2.9777	1.9251
3	3	1.5	2.9888	1.7661	2.9832	1.8956	2.9777	2.0240
4	2	.5	2.3222	1.4092	2.4832	1.6875	2.6443	1.9248
5	2	1.0	2.3222	1.5789	2.4832	1.7952	2.6443	1.9848
6	2	1.5	2.3222	1.8272	2.4832	1.9615	2.6443	2.0808
7	1	.5	1.6555	1.6358	1.9832	1.9099	2.3110	2.1067
8	1	1.0	1.6555	1.7842	1.9832	2.0056	2.3110	2.1616
9	1	1.5	1.6555	2.0072	1.9832	2.1558	2.3110	2.2502
10	0	.5	0.9888	1.9587	1.4832	2.2344	1.9777	2.3815
11	0	1.0	0.9888	2.0842	1.4832	2.3168	1.9777	2.4302
12	0	1.5	0.9888	2.2781	1.4832	2.4480	1.9777	2.5092
13	-1	.5	0.3222	2.3383	0.9832	2.6237	1.6443	2.7211
14	-1	1.0	0.3222	2.4444	0.9832	2.6942	1.6443	2.7638
15	-1	1.5	0.3222	2.6117	0.9832	2.8078	1.6443	2.8337
16	-2	.5	-.3445	2.7513	0.4832	3.0528	1.3110	3.1045
17	-2	1.0	-.3445	2.8420	0.4832	3.1136	1.3110	3.1420
18	-2	1.5	-.3445	2.9871	0.4832	3.2124	1.3110	3.2035

TABLE 4.26
MATH OUTCOMES, SENSITIVITY ANALYSIS
GRADES AS UNITS, COMBINED TREATMENTS

	prior		prior N = 24 posterior		prior N = 12 posterior		prior N = 6 posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.2846	1.9126	2.0461	2.1493	1.8554	2.3070
2	3	1.0	2.2846	2.0082	2.0461	2.2050	1.8554	2.3358
3	3	1.5	2.2846	2.1582	2.0461	2.2949	1.8554	2.3831
4	2	.5	1.7846	1.7810	1.7128	2.0449	1.6554	2.2367
5	2	1.0	1.7846	1.8834	1.7128	2.1034	1.6554	2.2665
6	2	1.5	1.7846	2.0425	1.7128	2.1974	1.6554	2.3152
7	1	.5	1.2846	1.7912	1.3795	2.0528	1.4554	2.2420
8	1	1.0	1.2846	1.8929	1.3795	2.1112	1.4554	2.2717
9	1	1.5	1.2846	2.0514	1.3795	2.2048	1.4554	2.3204
10	0	.5	0.7846	1.9407	1.0461	2.1720	1.2554	2.3224
11	0	1.0	0.7846	2.0350	1.0461	2.2271	1.2554	2.3510
12	0	1.5	0.7846	2.1831	1.0461	2.3161	1.2554	2.3980
13	-1	.5	0.2846	2.2073	0.7128	2.3856	1.0554	2.4704
14	-1	1.0	0.2846	2.2848	0.7128	2.4360	1.0554	2.4973
15	-1	1.5	0.2846	2.4178	0.7128	2.5176	1.0554	2.5416
16	-2	.5	-.2154	2.5391	0.3795	2.6712	0.8554	2.6749
17	-2	1.0	-.2154	2.6118	0.3795	2.7163	0.8554	2.6999
18	-2	1.5	-.2154	2.7289	0.3795	2.7897	0.8554	2.7409

and the data will always be less than the sample estimate obtained from the data. However, this is not true in general. The extra feature of the blending process is that it also takes into account the difference between the prior mean and the sample mean as contributing to the posterior estimate of the standard deviation (cf. Novick and Jackson, 1974, p.224). Thus, for example, when the prior judged mean is 3.0, which is rather far from the sample mean of -0.13, then the posterior standard deviation is larger than the sample standard deviation (1.82) for all combinations. In row 3 of Table 4.21, we see that if there is a judged standard deviation of 1.5, which is by far the closest to the sample estimate of 1.82, then the posterior estimate of the standard deviation is largest when the number of prior cases equals the number of data cases. This demonstrates the complex mixing that affects the estimated posterior standard deviation. The example here has the between-means portion weighted less when prior N = 6, and that property more than compensates for the fact that the pooling of the two within-samples estimates is being weighted more heavily by the sample.

In general, inspection of Table 4.21 indicates that the posterior mean can be affected substantially by the size of the prior sample, if the prior mean differs considerably from the data mean. If the prior mean is close to the data mean (in this instance where the prior mean equals zero), then the relative size of prior sample and data sample has a fairly small effect on the posterior mean. Here, these posteriors range from -.04 to -.09, an essentially trivial difference. Of course, this result is not surprising in view of the formula which determines the posterior mean. It is simply a weighted average of the prior mean and the data mean, each weighted by the number of cases on which it is based. When the two means are similar, the different weights make very little difference to the final judgeent about the mean.

Conversely, when the prior mean does differ considerably from the sample mean, the degree of confidence felt (as indicated by the number of case equivalents for the prior mean) is quite influential in determining the value of the posterior mean.

Thus, for example, a person with very optimistic and strongly held beliefs has a posterior mean of 1.96 for the impact, whereas a person identical except for having only weak confidence in his judgment has a posterior mean of 0.91. This is a substantial difference. It focuses attention on the need to learn more about factors determining the confidence felt by judges, and ways to increase the perceived relative influence of a sample where that would be appropriate. Conversely, the comparison between a confident pessimist and an unsure pessimist is reflected in the difference between the -1.38 posterior mean and the -0.75 posterior mean. Here the gap is smaller because both of these judges have their prior means somewhat closer to the data mean than do the optimists.

Perhaps the general point revealed by Table 4.21 is that, where the data results are indeed "grey" (mean near zero, and rather large standard deviation), then prior beliefs can be influential, especially if they differ considerably from the sample, and especially if they are held with high conviction. The Bayesian paradigm accepts unquestioningly the premise that prior and posterior means are indeed equal. To the extent they lie "too far" apart, the Bayesian machinery accommodates this by generating a larger estimate of the process standard deviation. This is contrary to the style of hypotheses testing research, in which the prior judgment would simply be rejected as not consistent with the data. We believe that, for many purposes, the Bayesian paradigm has more practical utility. By referring to the composite population of interest as a unit--but also making clear that the outcome data have high variability, the Bayesian tends to structure the knowledge from empirical studies as more relevant. Instead of a homogeneous but irrelevant statistical definition of the population, a Bayesian analysis constructs a variable population but one which does match the research or policy target.

The remaining tables in this set (Tables 4.22 through 4.26) are similar in layout to Table 4.21. They can be reviewed rather quickly here, since they can easily be studied in more detail by the interested reader.

The sensitivity analysis for the Reading Outcomes on the Parent-Teacher experiment is reported in Table 4.22. Here the same set of priors as in Table 4.21 was used to ascertain the sensitivity of the posterior distribution to the patterns of prior judgment. The results again indicate that where the prior judgment about the mean impact differs considerably from the data mean, the size of the sample judged equivalent to the prior judgment has a noticeable influence upon the posterior distribution. This is most clear for the extremely pessimistic prior mean (-2). If one holds such a prior judgment with greater confidence (i.e., feels it is based on a sample of size 24) as compared with the minimal confidence (a sample size equivalent of 6), then the posterior mean changes by about one full score unit, from -.96 to +.07.

Table 4.23 presents the results when the sensitivity analysis is performed upon the combined data from both the TO and the PT experiments. In view of the clear differences in the outcomes obtained in the data from the two comparisons, especially with respect to mathematics gains, this combining together is less useful than it would otherwise be.

For these reading outcomes, the Student's t value for their being different is 2.83, with 22 degrees of freedom, indicating a quite small probability that they are the same. Yet, there is some interest in examining this prospect nevertheless (despite its unlikely status based on the sample). If we thus make the assumption of a single mean impact operating here, then we may use the results shown in Table 4.23 to indicate the sensitivity of these results to differing prior judgments. This table, reflecting a data mean of 0.4870, a standard deviation of 2.1881, and 24 cases (not 12), indicates the effect of doubling the effective data case base. The wide range of prior means (from plus 3 through minus 2) is dampened considerably because of the greater sample size. For the prior judgment based on knowledge of one school, the range of posterior means shrinks to only about 2.5 if the prior is based on 24 cases also, to 1.7 if the prior is based on 12 cases, and to only 1.0 if the prior is based on 6 cases.

The results for the Teacher-Only treatment sensitivity analyses on the Math tests, shown in Table 4.24, are quite similar to those for the same comparison on the Reading test. The data mean for this impact on Math is 0.17, with a standard deviation of 1.83. The pattern of sensitivity resulting resembles closely the corresponding sensitivity analysis for Reading.

The sensitivity analysis for the Parent-Teacher comparison on Math is shown in Table 4.25. This result in the data was very strong in favor of a treatment impact, with the data mean being 2.97, and a standard deviation of 2.22, based on 12 cases as in the corresponding table (4.22) for Reading. Here the range of priors entertained was barely wide enough at the upper end to embrace the data. Thus, for all except those priors assuming a mean impact of three units, the incorporation of the prior tends to dampen the apparent effect. For example, if one believed as his prior judgment that the treatment will, on the average, make no difference, and that it will fluctuate with a standard deviation of 1.5, and that this judgment is equivalent in informative value to data from 12 cases, the implied posterior shows a mean impact of 1.48 units, with a standard deviation of 2.45. The implication here of perhaps greatest interest is in the standard deviation. Although the combination of the prior judgment and the data suggest a mean impact that may be regarded as moderate, they also imply a large fluctuation in the impacts to the expected. We shall return to this in a moment.

Table 4.26 completes this set. As with its counterpart for Reading (4.23), this table incorporates the premise that the outcomes from the TO comparisons and those from the PT comparisons can be combined as if drawn from a single population. This premise is quite unlikely, nearly impossible, on the basis of the data (_t_ value = 6.74 on 22 df). However, it is reported for the sake of completeness, and because once again it indicates the large standard deviations in the posterior, despite the small prior estimates.

This empirical finding of large standard deviations is perhaps the most ineresting feature of the data here. It is a result which does not emerge peculiarly from use of the Bayesian analysis, but rather is evident from an inspection of the data. However, the fact that _all_ of the observed standard deviations of the impact scores were larger than the largest of the standard deviations used in the sensitivity analysis as a prior judgment (1.5) indicates how far apart our intuition and this data are with respect to an expectation of relatively stable outcomes.

The implied message is in some respects a pessimistic one, but one that cannot be neglected. The clear implication is that variation between the units under consideration is large, and they also will react differently, perhaps in unpredictable patterns, to the introduction of a new feature to their operation. Two implications of this are noteworthy.

First, the analyses of outcomes of educational experiments probably should pay considerably more attention to the variability of reaction between schools, or between classrooms. It is naive and counterproductive to allow ourselves to pretend that an impact close to the average can be expected on each unit. The second implication is perhaps more in terms of philosophy. This wide range of response between units, greater than would be expected if those units were simple aggregates, is itself evidence for the real importance of school and subgroup structure. Just as in medicine and psychotherapy there is clear recognition that individuals do differ greatly in their capacities and responses, even though no theoretical dimension explains those differences very well, so too the social units which are reflected in this analysis differ one from another. The question of importance seems to us not to be how to account for such differences by a simple tidy theory, but rather how to identify in a clinical sense the particular pattern of response and capability in such a concrete group, and then to adapt the prescribed policy manipulation accordingly. Such a "clinical" paradigm has already become popular as an approach which teachers might use with individual students. The implication we draw from this analysis is that a parallel new awareness of,and sensitivity to, the individual classroom, or the individual school, is needed if policies at that level are to be developed more efficiently. Such an approach requires that a model be specified and estimated--even if the estimation is partly only judgmental. When this has been done, then the problem of designing or tailoring a policy to that model (like a clothesmaker's mannequin) can be approached systematically.

Since from many points of view the school rather than the classroom or grade is the appropriate unit of analysis here, we also carried out a sensitivity analysis on the same six comparisons, but treating each school as one case. This illustrates several points. First, from the point of view of a Federal or State agency implementing incentive schemes, it would seem unavoidable to use the school as the unit in which policies would be implemented. Thus, we find that alternate criteria or perspectives seem to imply that results aggregated to different levels are most appropriate.

Second, this choice of school as the unit of analysis reveals the potentially severe constraints of such a view, proper though it is. That is, despite the scope and cost of the study, we find that data is available for, say, the reading outcomes comparison on the TO treatment, on only two pairs of cases. Thus, we in fact use a Student's t distribution based on two cases and on one degree of freedom. In short, although approximately 80 teachers and 1800 students are involved, the comparison of schools is just barely possible.

If only one pair of schools had been observed, it would have been literally impossible to obtain an estimate of any kind regarding the standard deviation of impact. With the two pairs, such estimates of the standard deviation can be and are obtained. However, they cannot be regarded as very stable. Notice too that what is usually a minor point--whether to use the maximum likelihood estimator (modal) of the standard deviation (using N in the denominator)--now becomes quite important. That is, it reduces for this case to the choice of dividing by 2 or dividing by 1, thus making a very considerable difference to the subsequent interpretation. There simply is no good answer here; the paucity of data cannot be remedied by clever reasoning. We have chosen for incidental reasons to use the unbiased estimator. This was done (a) because it is more nearly in keeping with customary practice, and (b) because it results in larger estimates of the standard deviation. This leaning toward larger estimates at least will not allow us to feel falsely confident.

When these results based on very scanty data (though costly enough to collect and process) are combined with a range of alternative prior judgments about the impacts of the treatment, a pattern of posterior judgments is obtained. Table 4.27 shows such posterior judgment parameters, implying a Student's t, for a variety of prior judgments, and for the Reading outcomes on the Teacher-Only treatment comparison.

Table 4.27 reflects the sizeable sensitivity of results to even a prior judgment based on very few cases. For example, the range of posterior means in the columns involving a prior N = 2 is about 2.5 units, and the range in the columns where the prior N = 6 is about 3.75 units. It should be mentioned, of course, that since the units here are schools, an entirely different perspective is involved as far as the kinds of experience that would lead to a more confident

TABLE 4.27

READING OUTCOMES, SENSITIVITY ANALYSIS
SCHOOLS AS UNITS, TEACHER-ONLY TREATMENT

	prior		N = 6 posterior		N = 4 posterior		N = 2 posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.2225	1.6537	1.9633	1.8905	1.4450	2.2991
2	3	1.0	2.2225	1.8330	1.9633	2.0338	1.4450	2.3793
3	3	1.5	2.2225	2.0979	1.9633	2.2525	1.4450	2.5072
4	2	.5	1.4725	1.1957	1.2967	1.3543	0.9450	1.6359
5	2	1.0	1.4725	1.4334	1.2967	1.5480	0.9450	1.7467
6	2	1.5	1.4725	1.7596	1.2967	1.8259	0.9450	1.9173
7	1	.5	0.7225	0.7903	0.6300	0.8721	0.4450	1.0325
8	1	1.0	0.7225	1.1179	0.6300	1.1502	0.4450	1.2004
9	1	1.5	0.7225	1.5137	0.6300	1.5035	0.4450	1.4374
10	0	.5	-.0275	0.5653	-.0367	0.5950	-.0550	0.6752
11	0	1.0	-.0275	0.9719	-.0367	0.9574	-.0550	0.9116
12	0	1.5	-.0275	1.4093	-.0367	1.3617	-.0550	1.2066
13	-1	.5	-.7775	0.7174	-.7033	0.7835	-.5550	0.9198
14	-1	1.0	-.7775	1.0676	-.7033	1.0846	-.5550	1.1050
15	-1	1.5	-.7775	1.4770	-.7033	1.4540	-.5550	1.3586
16	-2	.5	-1.5275	1.0999	-1.3700	1.2412	-1.0550	1.4953
17	-2	1.0	-1.5275	1.3545	-1.3700	1.4502	-1.0550	1.6158
18	-2	1.5	-1.5275	1.6960	-1.3700	1.7438	-1.0550	1.7989

TABLE 4.28

READING OUTCOMES, SENSITIVITY ANALYSIS
SCHOOLS AS UNITS, PARENT-TEACHER TREATMENT

			N = 6		N = 4		N = 2	
	prior		posterior		posterior		posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.5650	1.0957	2.4200	1.2448	2.1300	1.5312
2	3	1.0	2.5650	1.3511	2.4200	1.4532	2.1300	1.6491
3	3	1.5	2.5650	1.6932	2.4200	1.7463	2.1300	1.8288
4	2	.5	1.8150	0.7618	1.7533	0.8502	1.6300	1.0510
5	2	1.0	1.8150	1.0979	1.7533	1.1337	1.6300	1.2163
6	2	1.5	1.8150	1.4990	1.7533	1.4909	1.6300	1.4507
7	1	.5	1.0650	0.6785	1.0867	0.7502	1.1300	0.9297
8	1	1.0	1.0650	1.0419	1.0867	1.0608	1.1300	1.1133
9	1	1.5	1.0650	1.4585	1.0867	1.4362	1.1300	1.3655
10	0	.5	0.3150	0.9167	0.4200	1.0342	0.6300	1.2745
11	0	1.0	0.3150	1.2106	0.4200	1.2775	0.6300	1.4140
12	0	1.5	0.3150	1.5834	0.4200	1.6029	0.6300	1.6200
13	-1	.5	-.4350	1.3117	-.2467	1.4976	0.1300	1.8397
14	-1	1.0	-.4350	1.5315	-.2467	1.6749	0.1300	1.9389
15	-1	1.5	-.4350	1.8404	-.2467	1.9346	0.1300	2.0939
16	-2	.5	-1.1850	1.7608	-.9133	2.0206	-.3700	2.4788
17	-2	1.0	-1.1850	1.9302	-.9133	2.1553	-.3700	2.5533
18	-2	1.5	-1.1850	2.1834	-.9133	2.3629	-.3700	2.6729

TABLE 4.29

READING OUTCOMES, SENSITIVITY ANALYSIS
SCHOOLS AS UNITS, COMBINED TREATMENTS

	prior		N = 6 posterior		N = 4 posterior		N = 2 posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.0300	1.5533	1.7875	1.6558	1.3833	1.7341
2	3	1.0	2.0300	1.6975	1.7875	1.7654	1.3833	1.7874
3	3	1.5	2.0300	1.9138	1.7875	1.9343	1.3833	1.8727
4	2	.5	1.4300	1.1215	1.2875	1.2076	1.0500	1.3129
5	2	1.0	1.4300	1.3139	1.2875	1.3540	1.0500	1.3826
6	2	1.5	1.4300	1.5836	1.2875	1.5679	1.0500	1.4912
7	1	.5	0.8300	0.8383	0.7875	0.9173	0.7167	1.0522
8	1	1.0	0.8300	1.0823	0.7875	1.1030	0.7167	1.1378
9	1	1.5	0.8300	1.3974	0.7875	1.3570	0.7167	1.2677
10	0	.5	0.2300	0.8647	0.2875	0.9442	0.3833	1.0758
11	0	1.0	0.2300	1.1029	0 2875	1.1254	0.3833	1.1596
12	0	1.5	0.2300	1.4134	0.2875	1.3753	0.3833	1.2873
13	-1	.5	-.3700	1.1801	-.2125	1.2681	0.0500	1.3689
14	-1	1.0	-.3700	1.3643	-.2125	1.4083	0.0500	1.4358
15	-1	1.5	-.3700	1.6256	-.2125	1.6150	0.0500	1.5408
16	-2	.5	-.9700	1.6241	-.7125	1.7296	-.2833	1.8047
17	-2	1.0	-.9700	1.7625	-.7125	1.8348	-.2833	1.8559
18	-2	1.5	-.9700	1.9717	-.7125	1.9979	-.2833	1.9383

TABLE 4.30

MATH OUTCOMES, SENSITIVITY ANALYSIS
SCHOOLS AS UNITS, TEACHER ONLY TREATMENT

	prior		N = 6 posterior		N = 4 posterior		N = 2 posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.3025	1.6539	2.0700	1.9114	1.6050	2.3999
2	3	1.0	2.3025	1.8331	2.0700	2.0533	1.6050	2.4767
3	3	1.5	2.3025	2.0980	2.0700	2.2701	1.6050	2.5999
4	2	.5	1.5525	1.2610	1.4033	1.4584	1.1050	1.8626
5	2	1.0	1.5525	1.4883	1.4033	1.6399	1.1050	1.9607
6	2	1.5	1.5525	1.8047	1.4033	1.9043	1.1050	2.1142
7	1	.5	0.8025	0.9722	0.7367	1.1255	0.6050	1.4763
8	1	1.0	0.8025	1.2531	0.7367	1.3525	0.6050	1.5983
9	1	1.5	0.8025	1.6161	0.7367	1.6633	0.6050	1.7831
10	0	.5	0.0525	0.8946	0.0700	1.0361	0.1050	1.3745
11	0	1.0	0.0525	1.1938	0.0700	1.2790	0.1050	1.5048
12	0	1.5	0.0525	1.5706	0.0700	1.6042	0.1050	1.6998
13	-1	.5	-.6975	1.0748	-.5967	1.2437	-.0395	1.6123
14	-1	1.0	-.6975	1.3343	-.5967	1.4523	-.0395	1.7247
15	-1	1.5	-.6975	1.6798	-.5967	1.7455	-.0395	1.8973
16	-2	.5	-1.4475	1.4178	-1.2633	1.6391	-.8950	2.0759
17	-2	1.0	-1.4475	1.6233	-1.2633	1.8026	-.8950	2.1644
18	-2	1.5	-1.4475	1.9175	-1.2633	2.0461	-.8950	2.3042

TABLE 4.31

MATH OUTCOMES, SENSITIVITY ANALYSIS
SCHOOLS AS UNITS, PARENT-TEACHER TREATMENT

			N = 6		N = 4		N = 2	
	prior		posterior		posterior		posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	3.1375	0.5660	3.1833	0.5855	3.2750	0.6206
2	3	1.0	3.1375	0.9722	3.1833	0.9514	3.2750	0.8719
3	3	1.5	3.1375	1.4096	3.1833	1.3575	3.2750	1.1770
4	2	.5	2.3875	0.9194	2.5167	1.0212	2.7750	1.1979
5	2	1.0	2.3875	1.2125	2.5167	1.2670	2.7750	1.3454
6	2	1.5	2.3875	1.5849	2.5167	1.5947	2.7750	1.5604
7	1	.5	1.6375	1.3675	1.8500	1.5523	2.2750	1.8669
8	1	1.0	1.6375	1.5797	1.8500	1.7239	2.2750	1.9647
9	1	1.5	1.6375	1.8807	1.8500	1.9772	2.2750	2.1178
10	0	.5	0.8875	1.8426	1.1833	2.1078	1.7750	2.5564
11	0	1.0	0.8875	2.0050	1.1833	2.2373	1.7750	2.6287
12	0	1.5	0.8875	2.2499	1.1833	2.4378	1.7750	2.7450
13	-1	.5	0.1375	2.3281	0.5167	2.6725	1.2750	3.2535
14	-1	1.0	0.1375	2.4587	0.5167	2.7759	1.2750	3.3106
15	-1	1.5	0.1375	2.6622	0.5167	2.9399	1.2750	3.4037
16	-2	.5	-.6125	2.8187	-.1500	3.2419	0.7750	3.9541
17	-2	1.0	-.6125	2.9275	-.1500	3.3275	0.7750	4.0012
18	-2	1.5	-.6125	3.1003	-.1500	3.4655	0.7750	4.0786

TABLE 4.32

MATH OUTCOMES, SENSITIVITY ANALYSIS
CHOOLS AS UNITS, COMBINED TREATMENTS

			N = 6		N = 4		N = 2	
	prior		posterior		posterior		posterior	
	μ	σ	μ	σ	μ	σ	μ	σ
1	3	.5	2.5520	1.5459	2.4400	1.7377	2.2533	2.0482
2	3	1.0	2.5520	1.6907	2.4400	1.8424	2.2533	2.0935
3	3	1.5	2.5520	1.9078	2.4400	2.0048	2.2533	2.1669
4	2	.5	1.9520	1.4205	1.9400	1.6144	1.9200	1.9447
5	2	1.0	1.9520	1.5769	1.9400	1.7266	1.9200	1.9923
6	2	1.5	1.9520	1.8077	1.9400	1.8990	1.9200	2.0693
7	1	.5	1.3520	1.4986	1.4400	1.6910	1.5867	2.0088
8	1	1.0	1.3520	1.6476	1.4400	1.7985	1.5867	2.0549
9	1	1.5	1.3520	1.8697	1.4400	1.9645	1.5867	2.1296
10	0	.5	0.7520	1.7533	0.9400	1.9441	1.2533	2.2260
11	0	1.0	0.7520	1.8821	0.9400	2.0383	1.2533	2.2677
12	0	1.5	0.7520	2.0794	0.9400	2.1862	1.2533	2.3357
13	-1	.5	0.1520	2.1217	0.4400	2.3165	0.9200	2.5578
14	-1	1.0	0.1520	2.2295	0.4400	2.3960	0.9200	2.5941
15	-1	1.5	0.1520	2.3983	0.4400	2.5232	0.9200	2.6537
16	-2	.5	-.4480	2.5553	-.0600	2.7604	0.5867	2.9657
17	-2	1.0	-.4480	2.6455	-.0600	2.8274	0.5867	2.9971
18	-2	1.5	-.4480	2.7892	-.0600	2.9359	0.5867	3.0488

or less confident prior judgment. Also, note that the sample estimate of the standard deviation for this table is .81, quite a lot smaller than found for the between-grade comparisons. However, this estimate is based on only two cases, and thus cannot be regarded as very stable.

In fact, the next table, Table 4.28, which reports the Reading outcomes for the School level analyses of the Parent-Teacher comparison, contains a somewhat larger estimated standard deviation based on the data, namely 1.19. Here the sample mean is clearly favorable to the treatment, although it is not large in magnitude.

All of these results, in the tables just discussed and in Tables 4.29 through 4.32, indicate that formalized prior judgments do substantially influence the interpretation made of school impact data, particularly when the unit of analysis is seen as being the grade or the school. Though perhaps somewhat frustrating to a researcher's sense of self-importance, they do tend to make clear that school policy formation workers are not necessarily behaving in an anti-scientific or irrational manner, when they discount heavily the efforts made to draw practical conclusions from the results of typical research studies.

5.0 Using Formal Models of Social Process

5.1 Introduction

As suggested in the previous chapters, especially Chapters 2 and 4, the Bayesian inference paradigm has, as one of its less obvious features, the property that it focuses an increased amount of attention on the need for a clearly specified formal model of the data-generating process. Using the Bayesian approach also indicates that such modeling can be both complex and difficult. It is complex in that the models must deal with the proximate data generating process (as indicated in Chapter 2) and with the more global context within which that process occurs. It is difficult because the phenomena are extremely elusive. Numerous decisions must be made as to whether the formalization will attempt to (1) model so as to reflect the full structure, or (2) substitute a "black-box" stochastic surrogate which produces judgmentally similar outputs but by a simplified process, or (3) hold as fixed the quantity in question, and thus make the model a conditional one. Such decisions must be made despite the vagueness and incompleteness of the available information as to the model's purposes, and as to the plausibility of various possible situational configurations. Thus, model specification is a highly problematic aspect of the research inference sequence for either style of analysis, but is more clearly evident to an analyst using the Bayesian framework.

The analysis work undertaken in this project on the O.E. Incentives Project data, which was described and reported in Chapter

4, has made this abstract point clear through a concrete example. We are consequently convinced that the major benefits of employing the Bayesian inference paradigm -- which can also be described as a paradigm stressing the idea of epistemic probability -- will be realized only when the data-generating models employed are made considerably more realistic than the highly simplified and abstract Normal distribution process now used typically. These simplified processes are in fact usually quite reasonable representations of the proximate process, but they neglect a whole range of interlocking circumstances and contingencies which affect the observed outcomes. Therefore, as is true in general, it seems clear that models for the impact of education programs on student achievement must be upgraded considerably.

The models to be used must be dynamic and synthetic (cf. Coleman, 1964). That is, they must involve time as an integral feature; they must be synthetic in the sense that they should begin with a basic process characterizing each individual entity at the most disaggregated level that needs to be considered, and must create other kinds of outcomes (more aggregate ones) as combinations and emergents from these ultimate beginnings. In education program analyses, therefore, this means that our models should begin with the individual student, and derive other consequences by aggregating individual student outcomes faithfully. The models used must also be operational ones, capable of actually generating batches of output if given input and allowed to run. They must embody causal patterns of relationship but also other non-causal dependencies as well. Finally, they must be implemented (as with Bayesian analysis programs) in a way that is flexible and powerful for use in actual research inference activities. These are very sizeable and difficult requirements. Our work in this project toward meeting them is little more than a groundbreaking, but even that is of value.

In this chapter, we describe and report on the work we have done thus far to delineate and develop a model for educational program impacts, and for certain other main issues of primary interest to policy-oriented educational research. The model has been formalized via a computer program, and is described in relation to that program. Stimulated in part by this development work on the model and program, and the future challenges involved in pursuing that line of attack on the inference problem, we describe also a strategy we have evolved for making best use of such models. The strategy is at present in a crude first stage, but is intended as a sensible and feasible general approach for using formal models systematically in conjunction with the Bayesian inference paradigm. By this joint use of formal models and Bayesian inference, it will be possible to consolidate and advance our scientific knowledge in a variety of social science areas. From a purely logical point of view, description of this general strategy should come first in order. However, it is pedagogically preferable to describe the concrete model and computer program first as an example, and then to illustrate the strategy by reference to the example.

5.2 The Classroom Dynamics Model - Background

We have worked in this project in some depth with the data from the O.E. Incentives Project, including some data that is not discussed in this Report. We have also a general familiarity with several other studies of compensatory education efforts done over the past decade. Also, we have detected a fairly clear trend since about 1972 in the kind of investigatory approach taken in studying compensatory education programs. All of these inputs of information point in the same general direction. Lately, a clear disenchantment has occurred among researchers with the kind of global, oversimple comparisons exemplified by the Project Head Start evaluation (Westinghouse/Ohio, 1969) and others. Instead, some of the best workers in the field have turned their attention to a more careful description of the educational experiences -- and patterns of experience -- that students have in school. A correlation study like the 1966 Coleman Report would be seen today as too gross to be of much value.

Researchers instead are again concerned with specific, short-term classroom processes; they seek to understand the ways in which these processes aggregate and combine to produce the larger distributional patterns of achievement and motivation. This new thrust is reflected quite clearly in a paper by Wiley and Harnischfeger (1975). and in a new book by William Cooley and Paul Lohnes (1976). Yet, the Wiley and Harnischfeger paper is disappointing. Despite a strongly and pleasantly flamboyant style, and an optimistic, positive, hortatory tone, the paper as a work of logical analysis and modeling is primitive and shallow. It fails, in short, to do much more than to whip up some enthusiasm. Such works are valuable, but only as catatysts to further action and much refinement. The book by Cooley and Lohnes (1976) is another matter. Although it too lacks nothing in tone, it accompanies that by much substance. The thrust of the Cooley and Lohnes work, however, is largely toward strategies of data reduction that can be truly feasible, and yet grounded in solid theory about learning processes. Both Wiley and Harnischfeger on the one hand, and Cooley and Lohnes on the other, cite the work of J. B. Carroll (1963) as seminal in the development of models for school learning processes.

Our work too draws on Carroll, but seeks to embed his ideas in a more applied and contemporary context. That is, one of the basic design criteria we adopted was that our model should include substantial detail on several aspects of school life that are seen as problematic in today's society. In this respect it is distinct from most other approaches. A second design criterion employed is that the model should not be a piecemeal or fragmentary one. We have noted in the earlier paragraphs of this chapter that educational researchers have once again begun to turn their attention toward the basic processes of school operation and classroom experiences. In a sense, this can be described as a parallel to the "back to basics" theme now cresting in popular discussions of education. Just as the consumers of education lately have renewed their interest in basic educational outputs, so the researchers and analysts of education have lately begun to look more closely at the basic or fundamental features of school life. As a reaction to the confusion generated by education studies in the 1960's, this tendency is a sound one. The danger is, however, that the new studies will be myopic and unintegrated. They may, in short, neglect the larger and more comlex system that is the actual school. This danger is understandable, because the total phenomenon is so very complex. But, in large measure it can be avoided by simply casting one's model in the form of a true dynamic, synthetic, structural simulation package. To design and use such a package will be costly, but not nearly as costly as it was a few years ago. The development of cheaper computing power has made available a resource which should be exploited.

5.3 The Classroom Dynamics Model - Specifics

These and similar considerations led us to conclude that improved research inference would be greatly facilitated by the availability of a rich but flexible simulator package which could unambiguously and completely operationalize the more global as well as the proximate aspects of our data-generating model. Accordingly, we decided to develop such a simulator. Our first explorations of the literature led us to consider the FEHR-Practicum package, developed by Lee Collet and others at The University of Michigan (1974), as a baseline. This package was attractive because descriptions of it indicated that it did provide true individual change in learning through (simulated) time, and according to a flexible set of parameters.

We acquired the package and, with the generous and gracious assistance of Dr. Collet and of Robert Parnes, the chief programmer on the FEHR project, spent some months studying its insides and its capabilities. Despite the enormous versatility of this package, and the realism and richness of its supporting materials, we eventually concluded that it could be of no real use for the purposes we had in mind. That is, the FEHR-Practicum package had itself been designed as a vehicle for teaching graduate students the ideas of educational research. Thus, although it did produce data that was quite realistic in appearance and basic properties, this production was accomplished through procedures which did not have any structural fidelity. In short, FEHR is a highly successful "black-box" simulator, but does not seek to be a structural modeling device.

This conclusion in turn led us to begin work on the development of a basic model and program of our own. The model was labeled the Classroom Dynamics Model, and a FORTRAN program embodying it has been written. A copy of that program, together with a sample file of input parameters, and a sample listing of output, is given in Appendix 5A. The model is at present in an intermediate stage of development but represents one plateau from which further development can be made. To attempt such development within the present project, however, would have been impossible.

The work of developing the Classroom Dynamics Model to its present state was guided by several general objectives. The work also highlighted several often latent problems in any effort at research inference, and caused us to evolve some additional heuristic principles, as well as setting some agenda items for future development work. In the remainder of this section, we shall discuss some of these aspects of the model.

The general inspiration for the Classroom Dynamics Model was drawn from the new book by Cooley and Lohnes (1976), and in particular their discussion of the work of Carroll (1963). But, in addition, a consciously chosen objective was to incorporate in the model representations of some of the principle problematic aspects of public school operation at present. Thus the model as presently implemented embodies several main themes:

(1) each student's instantaneous learning rate depends on his individual ability, his individual motivation, and his actual exposure to instruction;

(2) each student's instantaneous motivation level depends on his individual rate of progress in learning, on the amount he has learned, on the actual teacher time devoted to raising his motivation, and on peer influence;

(3) each student's attitude toward the teacher depends on the amount that student is learning, and on the amounts of teacher time with him (teaching and motivating/controlling);

(4) individual students develop likings for one another on the basis of sex, race, and similarity of learning and motivation patterns;

(5) the teacher allocates his/her time in a mixture of group instruction and individual contacts; a student's individual contact time depends on the student's perceived learning rates and motivation levels, and (potentially) on the teacher's liking for the student;

(6) absenteeism by students is partly dependent on recent school experience, and in turn influences subsequent school experience.

The simulation as presently programmed (see Appendix 5A) constructs one classroom of 21 students and one teacher. The actual programming langauage used is FORTRAN, but the logic of the program follows Jay Forrester's terminology as discussed in his book Principles of Systems (1968). Essentially, the simulator approximates a continuous-time, continuous-variable multi-equation system by advancing it through very small steps. The Classroom Dynamics Simulator has a simulated time span of one academic year (180 days or so). The simulator operates on each day and within a day on each of the students.

The sample output from the simulator, along with the particular configuration of input that produced it, indicates a rather unexceptional class situation. However, this is the result of rather lengthy work to tune the various parameters for such a result. This tuning process in our work to date was done in a cut-and-try manner, and thus is one area in which a need for more powerful procedures is quite apparent.

Notice that a great deal of output is produced if desired, but that the output can be rather easily controlled.

The output falls into several categories. Most broadly, there are options for graphed output and other options for tabular output. Graphed output (on the line printer) is available for five variables, 3 pertaining to student outcomes, and 2 pertaining to teacher behavior.

Graphs are available for the knowledge level of each student, the motivation level of each student, and the moving-average learning rate of each student. Each of these can be plotted with very flexible control of the layout. The two teacher time allocations (allocation of instructional time to each student, and allocation of "control" time to each student) likewise may be plotted on request.

The tabular output provides these same 5 quantities (appropriately scaled) for each student and according to flexible limits. In addition, summary statistics for the entire class at each time point (amount of teacher's time to group instruction, average amount of knowledge, average level of motivation, average rate of learning, and average amount of liking among students) are also provided. The liking matrices (showing attractions between students) are printed at several time points as requested, and the "kid-to-teacher" likings are also printed over time. This output is produced and stored in disk files as well as on paper, so it is readily subjected to further processing.

Many of the parameters governing the basic relationships in the model are read in from an input file, thus making it quite convenient to change them. Other constants and coefficients are incorporated in the actual program. This latter step was taken sometimes only as a temporary expedient during development, but in other cases is a deliberate choice because the necessary amount of flexibility can be gained without making those quantities variable. It should be mentioned that several of the particular formulas used to represent processes are chosen because they are mathematically convenient and reflect the basic ideas we have about the process. In a sense, these are our quasi-intentional rough approximations to the micro-dynamics of the process. At several points, chance events, usually following either a Normal distribution or a uniform distribution, may occur.

As a first result, the model and program can stand on their merits. They explicate a reasonable set of propositions about classrooms, and can be made to produce various reasonable patterns of output in response to various reasonable input quantities. In this sense the model passes one weak test of validity. At the same time, however, it also can be made to do strange things by the choice of other input mixtures. This latter property is an intriguing one, for we cannot be sure whether to regard it as suggesting validity or invalidity. The model does, however, serve very well as an explication device for the conjunction of the several basic processes described earlier.

Further work with the model lies in two directions. First, additional richness of mechanism should be introduced and compared with the present version. Most basically, we might try to convert to a "social actor" kind of model, in which each student and the teacher confronts at each time a mix of challenges and is given a set of resources. This approach would be developed by blending and operationalizing the broad ideas about collective action found in James Coleman's recent book on the subject (1973), and Walter Mischel's social learning conception of personality (1973). A second enrichment of the process model would be to explicate the contacts between teacher and student. These would include instructional contacts and "control" contacts. A third area for enrichment of the process model would be with respect to the development and influence of interpersonal bonds. Fourth and last, the mechanisms of learning and motivation which operate on the student outside of school (when at home, or when absent) might be introduced.

Clearly, this is a massive undertaking, and some priority setting must be done. Separate from these substantive enrichments, however, is the question of expansion and exploitation of such a model. The systematic variation through sets of parameter values needs to be programmed, as does the combination of one cohort of students through successive teachers, or of aggregation across teachers and students to the level of school. These are sizeable programming tasks and require substantial expertise in computer simulation. They will form the core of a project subsequent to this present one.

The Classroom Dynamics Model has already taught us plainly, however, one key lesson. Quite simply, we do not know enough to successfully build such models at present, and our interpretation of research results should be cognizant of that fact. In other words, just as the single major finding from the reanalysis of the O.E. Incentives Project data was that of substantial sensitivity to the choice of prior SCL so in this chapter the key finding is that there is equally strong sensitivity to the global model of the process.

To re-emphasize this point, the problem is not that of validating our model vis-a-vis our theory but rather that of validating our theory vis-a-vis the world. The simulation is merely the messenger bearing that sad piece of news.

5.4 Toward a Heuristic of Inference

In most general treatments of Bayesian statistical inference, including the treatment we provided in Chapter 2, it is taken for granted that there is a well-defined formula (a rule in the form of a mathematical expression) which, if given as a condition the value of certain parameters, can associate with a set of data points or data vectors a matching set of probabilities (or probability densities). Because these relationships are expressible via a rule, it is also the case that one can generate not only the function known as the probability function (or probability density function), but also the counterpart function known as the likelihood function. This latter function, given as a condition the values of a data point or data vector, can associate with a set of parameter values a matching set of likelihoods (or likelihood densities). This situation is a convenient one, since it provides such great economy of effort in dealing with the problem of inductive inference. In fact, it can be said that the situation is almost a necessary one if we are to be able to perform inductive inference.

There are several points here to be brought out. For example, if we are interested only in a finite set of distinct values for one parameter, and have observed a particular set of data, then a likelihood value can be calculated for each possible value of the parameter. The inference then can use the likelihood ratio of any pair of values as an aid to further steps, or perhaps can simply choose the parameter value which maximizes the likelihood as the "best" single estimate of the parameter. Correspondingly if several different ranges of parameter values are being considered, then the area of the space beneath the likelihood density curve for each of these intervals (or regions) can be calculated, and used in the inference. This latter approach is exactly the basis for what is one of the major virtues of the Bayesian approach, the "HDR." This stands for "highest density region", and denotes the narrowest possible region of parameter values embodying a specified fraction (e.g. 50%) of the total area under the likelihood curve.

But notice that these formulations, though highly useful, are also highly incomplete. That is, our inference does not in any of these cases give us direct assurance that the parameter takes on a particular value, nor that it lies in a particular region. We are given only probabilistic results of our inference. This point has been emphasized because it is quite important, and too often forgotten in applied inference -- that is, in scientific interpretation of results.

These ideas, though they leave us in the end still dealing in uncertain inferences, and having to state our conclusions in probabilistic terms, are reasonably clear as procedures. Also, they are readily made operational. The problem to which we wish to address ourselves in this section, however, is a slightly different one. We are concerned with the appropriate and/or feasible ways of extending, generalizing, and implementing this "likelihood inference" logic when the complete data-generating model under consideration is (a) complex, (b) multi-parameter, and (c) not analytic in form. The question we thus are discussing is nearly identical formally to the question of how to validate a computer simulation model properly. The only source of difference between the validation question and our more general inference question is that we are interested in not just practical-use models, but also scientific models, where even less external information about the system being modeled is available.

Three broad points are in order. We shall first state them baldly, then amplify each somewhat. The first main point about this whole question of complex likelihood inference is that it is inherent in the kinds of scientific topics we study. That is, it is not a defect introduced either by the Bayesian inference logic or by the use of computer simulation. The second main point is that within some limits we can attack these problems systematically, though only with a substantial investment of computing resources. The third main point, perhaps most important of all, is simply that the existence of many inference situations which are simply too complex and elusive to be dealt with in any adequate way should cause us to re-appraise our expectations and intuitions -- and our aspirations for the immediate future at least -- with respect to what "science" can tell us about complex social systems. In other words, we need to formulate, then to internalize, what might be described as an "uncertainty principle" of social science. Such a principle requires a narrow line to be defined, so that we neither seek the impossible, nor fail to stretch ourselves to greater advances in knowledge when these are possible.

The first point -- about the pervasiveness of the problem -- needs little explanation. Perhaps the only comment here is that the unpleasant implications of this point have contributed to its apparent novelty. But it is not a stranger, only one who has been suppressed and ignored. The exploration of the Bayesian inference approach, coupled with the utilization of computer modeling, in this project has made the point unescapably clear to us, and we believe it should now be faced squarely.

The second point, about the partial manageability of the implied inference problems, does bear some discussion. Basically, the conclusion seems to be that, with the aid of contemporary technology, we can deal with some of these problems, provided the level of complexity is kept down to a low or moderate level. The kinds of inference that can be used can only be heuristics, but they can nevertheless be quite powerful. We as a culture are probably at present just beginning the exploration of this kind of problem. By combining various elements of hard technology (computer power, data collection and reduction aids) and of intellectual advances in statistical methods (e.g. sequential analyses, multivariate discriminant analysis, experimental design, response surface methodology, etc.) and by applying them systematically to this problem of inference, there will be sizeable payoffs possible. These payoffs, however, will occur at least for the forseeable future only with a severely restricted range of problems.

This brings us to the third point. For the large class of problems not currently tractable by this new heuristic for inference, there must be some changes in outlook and in behavior by various interested parties. These changes will affect the researchers, the program designers, the managers, and the general public. They will occur not just within education but also within other social policy arenas. The changes will require work by philosophers to articulate the boundaries, by psychologists to study the human adaptive reactions, by social scientists and social designers to study and modify social patterns associated with the sense of knowledge versus ignorance.

These challenges are large in magnitude. They will not, however, disappear if ignored or denied. Rather, the problems that are in any case unavoidable will then be aggravated by the confusion and hostility that arise when frustration is recurrent. Thus, perhaps the most important conclusion of this Report -- and the one for which the Bayesian inference paradigm provides the intellectual key -- is that our scientific aspirations have at present outreached the grasp of our basic epistemology and research inference tradition. A revised paradigm is needed, and the elements for its construction are on hand. Work to define and develop that paradigm is critically needed.

```
123456789012345678901234567890123456789012345678901234567890123456789012

!                   CLASSROOM DYNAMICS MODEL
!                          JULY 1976

!                  PROGRAMMED BY JAY ROCKSTROH
!                  DESIGNED BY JAMES FENNESSEY

!                 DEPARTMENT OF SOCIAL RELATIONS
!                  THE JOHNS HOPKINS UNIVERSITY
!                      BALTIMORE, MARYLAND

!       A SIMULATION OF N (<=21) STUDENTS WITH EMPHASIS ON
!       LEARNING, MOTIVATION, FRIENDSHIP AND PEER PRESSURE,
!       AND WITH TEACHER TIME ALLOCATION AND ABSENCE AS FEATURES


      IMPLICIT REAL (A-Z)
      LOGICAL ABSENT,FLAG
      INTEGER LIK,LIKPER,ABTIME,DAYSAB,NUMPER
      INTEGER I,J,N,COUNT,DAY,NDAYS,PRT,PRTPER,PLT,PLTPER
      DIMENSION PLT1(5,55,21),FLAG(5)
      COMMON /COMM1/  IQLEV(21),LRNLEV(21),MOTLEV(21),LRNCHG(21),
     .         TCHTIM(21),TCHGOL(21),CONTIM(21),CONGOL(21),
     .         MOTCHG(21),RACE(21),SEX(21),CLASS,KT(21),KTG(21),TK(21),
     .         ABTIME(21),ABSENT(21),DAYSAB(21),LRNAVG(21),
     .         W11,W12,W13,W21,W22,W31,W32,W33,W34,W35,
     .         W41,W42,W43,W44,W51,W52,W53,W54,W55,W56,W61,W71,W72,
     .         W81,W82,W83,W84,
     .         DELAY1,DELAY2,DELAY3,DELAY4,DELAY5,DELAY6
      COMMON /COMM2/ LIKLEV(21,21),LIKGOL(21,21)
      COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN
      COMMON /COMM4/ AVGKNO,AVGMOT,AVGLRN,AVGLIK
      COMMON /COMM5/ DAY



      TYPE 10
   10 FORMAT(1H0,'NUMBER OF DAYS,PRINT INTERVAL,PLOT INTERVAL'/
     .   1H ,'PRINT INTERVAL FOR LIKING MATRIX,RANDOM SEED?')
      ACCEPT *, NDAYS,PRTPER,PLTPER,LIKPER,SEED
      COUNT=0
      PRT=PRTPER-1 ; PLT=PLTPER-1 ; LIK=LIKPER-1
      NUMPER=21 ! NUMBER OF KIDS PER PLOT


      CALL OPENF                  ! OPEN INPUT AND OUTPUT FILES
      CALL INPUT                  ! READ DATA FROM INPUT FILE
      CALL HEDING                 ! WRITE HEADINGS ON OUTPUT FILES
      CALL SETRAN(SEED)           ! INITIALIZE RANDOM NUMBER GENERATOR
      CALL SETFLG(FLAG)           ! SET PLOT FLAGS


! . . . RATES AND AUXILIARIES . . .
      DO 700 DAY=0,NDAYS

      TOTLRN=0 ; TOTKNO=0 ; TOTMOT=0 ; TOTLIK=0 ; XN=0
      DO 40 I=1,N
        IF (ABSENT(I)) GO TO 40
        XN=XN+1
        TOTLRN=TOTLRN+LRNAVG(I)
        TOTKNO=TOTKNO+LRNLEV(I)
        TOTMOT=TOTMOT+MOTLEV(I)
        DO 40 J=1,N
          IF (ABSENT(J)) GO TO 40
          TOTLIK=TOTLIK+LIKLEV(I,J)
   40 CONTINUE
      AVGLRN=TOTLRN/XN
      AVGKNO=TOTKNO/XN
      AVGMOT=TOTMOT/XN
      AVGLIK=TOTLIK/(XN*(XN-1))

      DO 50 I=1,N
        ABTIME(I)=ABTIME(I)-1
        IF (ABTIME(I)>0) GO TO 50
        ABSENT(I)=.NOT.ABSENT(I)
        IF (ABSENT(I))
     *       ABTIME(I)=(4+8*(1-MOTLEV(I)))*RAN(0)**2+1
        IF (.NOT.ABSENT(I))
     *       ABTIME(I)=NORMAL(15.+15*(MOTLEV(I)-1),16.)
   50 CONTINUE

      J=N
      DO 60 I=1,N
        IF (ABSENT(I)) DAYSAB(I)=DAYSAB(I)+1
   60 CONTINUE


      TOTAL=0
      DO 90 I=1,N
        IF (ABSENT(I)) GO TO 80
        BIAS=TK(I)+W71*(RACE(I)-1.5)+W72*(SEX(I)-1.5)
        TCHGOL(I)=W11 + W12 * (LRNAVG(I)-AVGLRN)
     *                + W13 * (LRNLEV(I)-AVGKNO) + BIAS
        CONGOL(I)=W21 + W22 * (MOTLEV(I)-AVGMOT)
        IF (TCHGOL(I)<0) TCHGOL(I)=0
        IF (CONGOL(I)<0) CONGOL(I)=0
        GO TO 90
   80   TCHGOL(I)=W81+W82*TCHGOL(I) ; CONGOL(I)=W83+W64*CONGOL(I)
   90   TOTAL=TOTAL+TCHGOL(I)+CONGOL(I)
      DO 100 I=1,N
        TCHGOL(I)=TCHGOL(I)/TOTAL
  100   CONGOL(I)=CONGOL(I)/TOTAL

      TOTAL=0
      DO 110 I=1,N
  110   IF (.NOT.ABSENT(I)) TOTAL=TOTAL+CONTIM(I)+TCHTIM(I)

      DO 160 I=1,N
        IF (ABSENT(I)) GO TO 140
        MOTCHG(I)=W31 * (LRNAVG(I)-AVGLRN)
     *                + W32 * CONTIM(I)*(1-CLASS)/TOTAL
     *                + W33 * (LRNLEV(I)-AVGKNO)/AVGKNO
     *                + W34 * KT(I)
     *                + W35 * (RAN(0)-.5)
        DO 120 J=1,N
          IF (ABSENT(J)) GO TO 120
        INFLU=LIKLEV(I,J)
        IF (INFLU<.30) INFLU=0
        MOTCHG(I)=MOTCHG(I)+(MOTLEV(J)-MOTLEV(I))*INFLU/20.
 120 CONTINUE
      MOTCHG(I)=MOTCHG(I) * 4 * (MOTLEV(I)-.5) * (1.5-MOTLEV(I))

      LRNCHG(I)=W41 * ( IQLEV(I) -W42)
   *                  * (MOTLEV(I)-W43)
   *                  * ( CLASS * W44*TCHTIM(I)*(1-CLASS)/TOTAL )

      KTG(I)= .7*KTG(I)
   *              + .3*(240*LRNAVG(I)-.8)
   *              + (1-CLASS)*TCHTIM(I)*10*(RAN(0)-.2)
   *              + (1-CLASS)*CONTIM(I)* 5*(RAN(0)-.5)
        IF (KTG(I)>2) KTG(I)=2
        IF (KTG(I)<-2) KTG(I)=-2

      DO 130 J=1,N
        IF (J==I .OR. ABSENT(J) ) GO TO 130
        DISSIM=W51 + (LRNLEV(I)-LRNLEV(J))**2/AVGKNO *W52
   *                + (MOTLEV(I)-MOTLEV(J))**2 *W53
   *                + ABS(RACE(I)-RACE(J)) *W54
   *                + ABS(SEX(I)-SEX(J)) *W55
   *                + (LRNAVG(I)-LRNAVG(J))**2 *W56
        LIKGOL(I,J)=W61/(DISSIM*DISSIM*DISSIM)
 130  CONTINUE
      GO TO 160

 140  MOTCHG(I)=0 ; LRNCHG(I)=0 ; KTG(I)=0
      DO 150 J=1,N
        LIKGOL(J,I)=0
 150    LIKGOL(I,J)=0

 160 CONTINUE

     DO 180 I=1,N
       IF (ABSENT(I)) GO TO 180
       TOTAL=0
       DO 170 J=1,N
 170     TOTAL=TOTAL+LIKGOL(I,J)
       TOTAL=TOTAL/3.5
       DO 180 J=1,N
         LIKGOL(I,J)=LIKGOL(I,J)/TOTAL
 180 CONTINUE


! . . . PRINT ROUTINE . . .
     PRT=PRT+1
     IF (PRT#PRTPER) GO TO 500
     PRT=0 ; CALL OUT1

! . . . PRINT LIKING MATRIX . . .
 500 LIK=LIK+1
     IF (LIK#LIKPER) GO TO 550
     LIK=0 ; CALL OUTLIK

! . . . SAVE POINTS FOR PLOT . . .
 550 PLT=PLT+1
     IF (PLT#PLTPER .OR. COUNT>=55) GO TO 600
     PLT=0 ; COUNT=COUNT+1
     DO 560 I=1,N
       PLT1(1,COUNT,I)=LRNLEV(I)
       PLT1(2,COUNT,I)=MOTLEV(I)
       PLT1(3,COUNT,I)=LRNAVG(I)
       PLT1(4,COUNT,I)=TCHTIM(I)
560    PLT1(5,COUNT,I)=CONTIM(I)


! . . . UPDATE LEVELS . . .
600  CLASS=CLASS+(AVGLRN-.005)/DELAY6
     IF (CLASS>1) CLASS=1
     IF (CLASS<0) CLASS=0
     DO 700 I=1,N
       TCHTIM(I)=TCHTIM(I)+(TCHGOL(I)-TCHTIM(I))/DELAY1
       CONTIM(I)=CONTIM(I)+(CONGOL(I)-CONTIM(I))/DELAY2
       MOTLEV(I)=MOTLEV(I)+MOTCHG(I)
       LRNLEV(I)=LRNLEV(I)+LRNCHG(I)*LRNLEV(I)
       KT(I)=KT(I)+(KTG(I)-KT(I))/DELAY3
       LRNAVG(I)=LRNAVG(I)+(LRNCHG(I)-LRNAVG(I))/DELAY4
       DO 700 J=1,N
         IF (J==I) GO TO 700
         LIKLEV(I,J)=LIKLEV(I,J)+(LIKGOL(I,J)-LIKLEV(I,J))/DELAY5
700  CONTINUE

      CALL OUTAB(DAYSAB)

     CALL PREPLT(N,COUNT,PLT1,PLTPER,FLAG,NUMPER)

     END
```

```
      SUBROUTINE OPENF

      DOUBLE PRECISION FILNAM

      DO 10 I=20,28
        OPEN(UNIT=I,DEVICE='DSK',ACCESS='SEQOUT')
   10 CONTINUE
      TYPE 20
   20 FORMAT(1H0,'TYPE NAME OF INPUT DATA FILE')
      ACCEPT 15, FILNAM
   15 FORMAT(A10)
      OPEN(UNIT=2,DEVICE='DSK',ACCESS='SEQIN',FILE=FILNAM)

      RETURN
      END


      SUBROUTINE INPUT

      IMPLICIT REAL (A-Z)
      LOGICAL ABSENT
      INTEGER I,J,N,SYM,ABTIME,DAYSAB
      DOUBLE PRECISION SS
      DIMENSION RS(2),SS(2)
      COMMON /COMM1/ IQLEV(21),LRNLEV(21),MOTLEV(21),LRNCHG(21),
     .          TCHTIM(21),TCHGOL(21),CONTIM(21),CONGOL(21),
     .          MOTCHG(21),RACE(21),SEX(21),CLASS,KT(21),KTG(21),TK(21),
     .          ABTIME(21),ABSENT(21),DAYSAB(21),LRNAVG(21),
     .          W11,W12,W13,W21,W22,W31,W32,W33,W34,W35,
     .          W41,W42,W43,W44,W51,W52,W53,W54,W55,W56,W61,W71,W72,
     .          W81,W82,W83,W84,
     .          DELAY1,DELAY2,DELAY3,DELAY4,DELAY5,DELAY6
      COMMON /COMM2/ LIKLEV(21,21),LIKGOL(21,21)
      COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN
      DATA RS(1)/'WHITE'/
      DATA RS(2)/'BLACK'/
      DATA SS(1)/'   MALE   '/
      DATA SS(2)/'  FEMALE  '/
      DATA (SYM(I),I=1,21)/'ANDY','BEN','CINDY','DIANE','EGOR',
     . 'FRAN','GREG','HENRY','IONA','JASON','KELLI','LARRY','MARY',
     . 'NICK','OLIVE','PETER','QUINT','RITA','SANDY','TERRY','UXTON'/

      READ(2,*) N
      XN=N
      READ(2,11) DUMMY   !SKIP HEADER CARD
   11 FORMAT(A1)

      DO 10 I=1,N
        READ(2,*) LRNLEV(I),RACE(I),SEX(I),IQLEV(I),
     .          MOTLEV(I),KTG(I),TK(I)
        KT(I)=KTG(I)
        LRNCHG(I)=IQLEV(I)*.00005*MOTLEV(I)
        LRNAVG(I)=LRNCHG(I)
        TCHTIM(I)=1/XN
        CONTIM(I)=0
        ABTIME(I)=-10*ALOG(RAN(0))+1
        ABSENT(I)=.FALSE.       ! ASSUMES ALL PRESENT ON FIRST DAY
        DAYSAB(I)=0
        DO 10 J=1,N
   10     LIKLEV(I,J)=0
```

```
      READ(2,*) CLASS
      READ(2,*) W11,W12,W13
      READ(2,*) W21,W22
      READ(2,*) W31,W32,W33,W34,W35
      READ(2,*) W41,W42,W43,W44
      READ(2,*) W51,W52,W53,W54,W55,W56
      READ(2,*) W61
      READ(2,*) W71,W72
      READ(2,*) W81,W82,W83,W84
      READ(2,*) DELAY1,DELAY2,DELAY3,DELAY4,DELAY5,DELAY6

      CALL DATE(TDATE)
      CALL TIME(DAYTIM)
      WRITE(20,39) (TDATE(I),I=1,2),DAYTIM
   39 FORMAT(1H1,'CLASSROOM DYNAMICS MODEL'/1H0,2A5,5X,A5/1H/,
     . 'STUDENT INFORMATION'//1H0,'STUDENT     KNOWLEDGE        RACE',
     . '      SEX          IQ       MOTIVATION')
      DO 41 I=1,N
        WRITE(20,40) SYM(I),LRNLEV(I),RS(RACE(I)),SS(SEX(I)),
     .                      IQLEV(I),MOTLEV(I)
   40   FORMAT(1H0,X,A5,F11.0,8X,A5,X,A10,F7.0,F12.2)
   41 CONTINUE
      WRITE(20,42) W11,W12,W13,W21,W22,W31,W32,W33,W34,W35,
     .   W41,W42,W43,W44,W51,W52,W53,W54,W55,W56,W61,W71,W72,
     .   W81,W82,W83,W84
   42 FORMAT(1H/,'WEIGHTS'//1H ,3F12.6/1H ,2F12.6/1H ,5F12.6/1H ,4F12.6/
     .   1H ,6F12.6/1H ,F12.6/1H ,2F12.6/1H ,4F12.6)
      WRITE(20,43) DELAY1,DELAY2,DELAY3,DELAY4,DELAY5,DELAY6
   43 FORMAT(1H0,'DELAYS'//1H ,6F6.1)

      RETURN
      END



      SUBROUTINE HEDING

      IMPLICIT REAL (A-Z)
      INTEGER I,J,N,SYM
      COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN

      WRITE(21,21) (TDATE(J),J=1,2),DAYTIM
      WRITE(22,22) (TDATE(J),J=1,2),DAYTIM
      WRITE(23,23) (TDATE(J),J=1,2),DAYTIM
      WRITE(24,24) (TDATE(J),J=1,2),DAYTIM
      WRITE(25,25) (TDATE(J),J=1,2),DAYTIM
      WRITE(26,26) (TDATE(J),J=1,2),DAYTIM

   21 FORMAT(1H1,'DAY',4X,46('-'),' (SCALED) KNOWLEDGE',
     .  41('-'),X,2A5,2X,A5)
   22 FORMAT(1H1,'DAY',4X,46('-'),' (SCALED) MOTIVATION',
     .  40('-'),X,2A5,2X,A5)
   23 FORMAT(1H1,'DAY',4X,46('-'),' (SCALED) LEARNING RATE',
     .  37('-'),X,2A5,2X,A5)
   24 FORMAT(1H1,'DAY',4X,46('-'),' (SCALED) TEACHING TIME',
     .  37('-'),X,2A5,2X,A5)
   25 FORMAT(1H1,'DAY',4X,46('-'),' (SCALED) CONTROL TIME',
     .  38('-'),X,2A5,2X,A5)
   26 FORMAT(1H1,'DAY         CLASS     AVGKNO     AVGMOT     AVGLRN',
     . '     AVGLIK',10X,2A5,2X,A5)
   WRITE(28,28) (TDATE(I),I=1,2),DAYTIM
28 FORMAT(1H1,'DAY',4X,46('-'),' (SCALED) KID TO TEACHER',
  .  36('-'),X,2A5,2X,A5)

   DO 30 I=21,25
30   WRITE(I.49) (SYM(J),J=1,N)
   WRITE(28,49) (SYM(I),I=1,N)
49 FORMAT(1H ,6X,21(X,A5))

   RETURN
   END



   SUBROUTINE SETFLG(FLAG)

   LOGICAL FLAG
   DIMENSION FLAG(5)

   DO 10 I=1,5
10   FLAG(I)=.TRUE.

20 FORMAT(A5)

   WRITE(5,30)
30 FORMAT(1H0,' DO YOU WANT THE FOLLOWING PLOTS?  (YES OR NO)'///)

   WRITE(5,31)
31 FORMAT(1H+,'LEARNING LEVEL? ',$)
   READ(5,20) ANSWER
   IF (ANSWER=='NO') FLAG(1)=.FALSE.

   WRITE(5,32)
32 FORMAT(1H+,'MOTIVATION LEVEL? ',$)
   READ(5,20) ANSWER
   IF (ANSWER=='NO') FLAG(2)=.FALSE.

   WRITE(5,33)
33 FORMAT(1H+,'STUDENTS AVERAGE LEARNING RATE? ',$)
   READ(5,20) ANSWER
   IF (ANSWER=='NO') FLAG(3)=.FALSE.

   WRITE(5,34)
34 FORMAT(1H+,'TEACHING TIME? ',$)
   READ(5,20) ANSWER
   IF (ANSWER=='NO') FLAG(4)=.FALSE.

   WRITE(5,35)
35 FORMAT(1H+,'CONTROL TIME? ',$)
   READ(5,20) ANSWER
   IF (ANSWER=='NO') FLAG(5)=.FALSE.

   RETURN
   END


   REAL FUNCTION NORMAL(MEAN,VAR)
```

```
      IMPLICIT REAL (A-Z)

   10 U1=2*(RAN(0)-.5)
      U2=2*(RAN(0)-.5)
      S=U1**2+U2**2
      IF (S>1) GO TO 10

      N1=U1*SQRT(-2*ALOG(S)/S)
      NORMAL= N1 * SQRT(VAR) + MEAN

      RETURN
      END



      SUBROUTINE OUT1

      IMPLICIT REAL (A-Z)
      INTEGER I,J,N,SYM,ABTIME,ABSENT,DAYSAB,VECTOR,DAY
      DIMENSION VECTOR(21)
      COMMON /COMM1/ IQLEV(21),LRNLEV(21),MOTLEV(21),LRNCHG(21),
     *          TCHTIM(21),TCHGOL(21),CONTIM(21),CONGOL(21),
     *          MOTCHG(21),PACE(21),SEX(21),CLASS,KT(21),KTG(21),TK(21),
     *          ABTIME(21),ABSENT(21),DAYSAB(21),LRNAVG(21),
     *          W11,W12,W13,W21,W22,W31,W32,W33,W34,W35,
     *          W41,W42,W43,W44,W51,W52,W53,W54,W55,W56,W61,W71,W72,
     *          W81,W82,W83,W84,
     *          DELAY1,DELAY2,DELAY3,DELAY4,DELAY5,DELAY6
      COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN
      COMMON /COMM4/ AVGKNO,AVGMOT,AVGLRN,AVGLIK
      COMMON /COMM5/ DAY

      DO 8 I=1,N
    8   VECTOR(I)=LRNLEV(I)*10
      WRITE(21,100) DAY,(VECTOP(I),I=1,N)

      DO 9 I=1,N
    9   VECTOR(I)=MOTLEV(I)*100
      WRITE(22,100) DAY,(VECTOR(I),I=1,N)

      DO 10 I=1,N
   10   VECTOR(I)=LRNAVG(I)*10000
      WRITE(23,100) DAY,(VECTOR(I),I=1,N)

      DO 11 I=1,N
   11   VECTOR(I)=TCHTIM(I)*1000
      WRITE(24,100) DAY,(VECTOR(I),I=1,N)

      DO 12 I=1,N
   12   VECTOR(I)=CONTIM(I)*1000
      WRITE(25,100) DAY,(VECTOP(I),I=1,N)

      DO 13 I=1,N
   13   VECTOR(I)=KT(I)*1000
      WRITE(28,100) DAY,(VECTOR(I),I=1,N)

  100 FORMAT(1H ,I3,2X,21I6)

      WRITE(26,26) DAY,CLASS,AVGKNO,AVGMOT,AVGLRN,AVGLIK
   26 FORMAT(1H ,I3,2X,5F10.3)
```

```
      RETURN
      END


      SUBROUTINE OUTLIK

      IMPLICIT REAL (A-Z)
      INTEGER I,J,N,SYM,VECTOR,DAY
      DIMENSION VECTOR(21)
      COMMON /COMM2/ LIKLEV(21,21),LIKGOL(21,21)
      COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN
      COMMON /COMM5/ DAY

      IF (DAY==0) GO TO 40
      WRITE(27,10) DAY,(TDATE(I),I=1,2),DAYTIM
   10 FORMAT(1H1,'LIKING MATRIX     DAY=',I3,8X,2A5,5X,A5////)

      WRITE(27,15) (SYM(I),I=1,N)
   15 FORMAT(1H0,6X,21(X,A5))
      DO 30 I=1,N
        DO 16 J=1,N
          VECTOR(J)=1000*LIKLEV(I,J)
   16     IF (VECTOR(J)<300) VECTOR(J)=0
        WRITE(27,20) SYM(I), (VECTOR(J),J=1,N)
   20   FORMAT(1H0,A5,X,21I6)
   30 CONTINUE

   40 RETURN
      END



      SUBROUTINE OUTAB(DAYSAB)

      IMPLICIT REAL (A-Z)
      INTEGER DAYSAB,I,J,N,SYM
      DIMENSION DAYSAB(21)
      COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN

      WRITE(26,10) (SYM(I),I=1,N)
   10 FORMAT(1H0,5X,'DAYS ABSENT'/1H0,6X,21A6)
      WRITE(26,20) (DAYSAB(I),I=1,N)
   20 FORMAT(1H ,5X,21I6)

      RETURN
      END



      SUBROUTINE PREPLT(N.COUNT.A.PLTPER.FLAG.NUMPER)

      IMPLICIT REAL (A-Z)
      LOGICAL FLAG
      INTEGER I,J,N,N1,COUNT,K,NUMPER
      DIMENSION A(5,55,21),E(55,21),NAME(5),FLAG(5)

      DO 100 I=1,5
        READ(2,10) (NAME(J),J=1,5),MIN,MAX
```

```
 10   FORMAT(5A5,7X,2G)
      IF (.NOT.FLAG(I)) GO TO 100
      N1=N ; J=0
 20   IF (N1<=NUMPER) GO TO 30
      DO 25 K=1,NUMPER
        DO 25 L=1,COUNT
 25       E(L,K)=A(I,L,K+J)
      CALL PLOT(NUMPER,COUNT,E,J,NAME,MIN,MAX,PLTPER)
      J=J+NUMPER ; N1=N1-NUMPER
      GO TO 20

 30   DO 35 K=1,N1
        DO 35 L=1,COUNT
 35       E(L,K)=A(I,L,K+J)
      CALL PLOT(N1,COUNT,E,J,NAME,MIN,MAX,PLTPER)

100 CONTINUE

    RETURN
    END



    SUBROUTINE PLOT(NUMVAR,POINTS,A,ISYM,NAME,MIN,MAX,PLTPER)

    IMPLICIT REAL (A-Z)
    INTEGER I,J,K,N,SYM,PLOTV,PLTPER,DAY
    INTEGER ISYM,BLANK,PERIOD,DASH,NUMVAR,POINTS
    COMMON /COMM3/ TDATE(2),DAYTIM,SYM(21),N,XN
      DIMENSION NAME(5),A(55,21)
      DIMENSION PLOTV(100),SCALE(6)

      WRITE(20,40)
 40   FORMAT (1H1)
      WRITE(20,65) (NAME(J),J=1,5),(TDATE(I),I=1,2),DAYTIM
 65   FORMAT (1H ,5A5,X,2A5,2X,A5//)
      DO 69 J=1,6
        X=J
 69     SCALE(J)=((X-1.)/5.)*MAX + ((6.-X)/5.)*MIN
      WRITE(20,68) (SCALE(J),J=1,6)
 68   FORMAT(1H ,5X,6(4X,F10.4,5X)//)

      MULT=100./(MAX-MIN)
      DO 90 I=1,POINTS
        DO 72 K=1,100
          PLOTV(K)=' '
          IF (I-1==10*((I-1)/10)) PLOTV(K)='-'
 72       IF (K==1 .OR. K==(K/20)*20 ) PLOTV(K)='.'
        DO 73 K=1,NUMVAR
          J=(A(I,K)-MIN)*MULT+.5
          IF ( J<1 )  J=1
          IF ( J>100 )  J=100
 73       PLOTV(J)=SYM(K+ISYM)
        DAY=(I-1)*PLTPER
        WRITE(20,74) DAY,(PLOTV(J),J=1,100)
 74     FORMAT(1H ,5X,I6,4X,100A1)
 90   CONTINUE
      RETURN
      END
```

CLASSROOM DYNAMICS PROGRAM DOCUMENTATION

MODEL VARIABLES

ARRAYS

IQLEV	INTELLIGENCE QUOTIENT OF INDIVIDUAL
LRNLEV	ACCUMULATED KNOWLEDGE (LEARNING)
MOTLEV	LEVEL OF MOTIVATION
LRNCHG	LEARNING (RATE OF CHANGE IN LRNLEV)
TCHTIM	% OF NON-GROUP TIME ALLOTTED TO INDIVIDUAL FOR TEACHING PURPOSES
TCHGOL	DESIRED TCHTIM
CONTIM	% OF NON-GROUP TIME ALLOTTED TO INDIVIDUAL FOR BEHAVIOR CONTROL PURPOSES
CONGOL	DESIRED CONTIM
MOTCHG	MOTIVATION CHANGE (RATE OF CHANGE OF MOTLEV)
RACE	RACE OF INDIVIDUAL
SEX	SEX OF INDIVIDUAL
KT	KID TO TEACHER LIKING (STRENGTH OF INDIVIDUAL'S LIKING FOR TEACHER)
KTG	KID TO TEACHER GOAL (DESIRED KT)
TK	TEACHER TO KID LIKING (STRENGTH OF TEACHER'S LIKING FOR INDIVIDUAL)
ABTIME	DURATION OF CURRENT STATE OF ABSENCE OR PRESENCE
ABSENT	=0 IF INDIVIDUAL IS ABSENT =1 IF INDIVIDUAL IS PRESENT
DAYSAB	# OF DAYS INDIVIDUAL HAS BEEN ABSENT
LRNAVG	EXPONENTIALLY WEIGHTED AVERAGE OF INDIVIDUAL'S LEARNING RATE (LRNCHG)
LIKLEV	LIKING OF INDIVIDUAL I FOR INDIVIDUAL J
LIKGOL	LIKING GOAL (DESIRED LIKLEV)

SCALARS

DAY	MODEL TIME IN DAYS FROM START OF ACADEMIC YEAR
TOTLRN	SUM OF LEARNING RATES OF PRESENT STUDENTS
TOTKNO	SUM OF LEARNING LEVELS OF PRESENT STUDENTS
TOTMOT	SUM OF MOTIVATION LEVELS OF PRESENT STUDENTS
TOTLIK	SUM OF LIKINGS OF PRESENT STUDENTS
N	NUMBER OF STUDENTS
XN	NUMBER OF STUDENTS PRESENT
AVGLRN	AVERAGE LEARNING RATE OF PRESENT STUDENTS
AVGKNO	AVERAGE LEARNING LEVEL OF PRESENT STUDENTS
AVGMOT	AVERAGE MOTIVATION LEVEL OF PRESENT STUDENTS
AVGLIK	AVERAGE LIKING LEVEL OF PRESENT STUDENTS
TOTAL	LOCAL SUMMING VARIABLE
CLASS	% OF TIME SPENT TEACHING STUDENTS AS GROUP
BIAS	TEACHERS PREFERENCE FOR STUDENT
RAN(0)	RANDOM NUMBER (UNIFORM DISTRIBUTION)
INFLU	INFLUENCE OF ONE STUDENT ON ANOTHER (PEER PRESSURE)
DISSIM	DISSIMILARITY OF TWO STUDENTS
DELAY1	DELAY IN ADJUSTING TCHTIM
DELAY2	DELAY IN ADJUSTING CONTIM

DELAY3	DELAY IN ADJUSTING KT
DELAY4	DELAY IN ADJUSTING LRNAVG
DELAY5	DELAY IN ADJUSTING LIKLEV
DELAY6	DELAY IN ADJUSTING CLASS

sample of input cards for
Classroom Dynamics Program

```
21                          NUMBER OF STUDENTS
KNOW RACE  SEX    IQ  MOTIVATION  KT   TK
11    1     1    100     1.00    .40   0
 8    2     1    120     1.20    .40   0
12    1     2     97     1.07    .40   0
 7    2     2    103      .90    .40   0
10    1     1     90     1.02    .40   0
 7    1     2     95      .78    .40   0
13    2     1     85      .93    .40   0
10    1     1    100     1.03    .40   0
10    1     2    100     1.00    .40   0
10    2     1    101      .80    .40   0
10    2     2    107      .86    .40   0
 9    1     1     93     1.10    .40   0
 9    1     2    100     1.05    .40   0
 9    2     1     92      .90    .40   0
 9    2     2    100      .95    .40   0
11    1     1    105     1.12    .40   0
12    2     1    110     1.04    .40   0
10    1     2    115     1.17    .40   0
 9    1     2     78      .84    .40   0
11    2     2    108     1.14    .40   0
 8    1     1     86      .98    .40   0
 .9                 INITIAL CLASS TIME
  .2       -1.8       -.035                                  ! EQ. 1
  .1       -1.5                                              ! EQ. 2
  .05        .075      .0025    .0015      .005              ! EQ. 3
  .000055  0           .10     10.0                          ! EQ. 4
 1.0         .25      1.0       .15         .3     60.0      ! EQ. 5
 1.0                                                         ! EQ. 6
  0        0                                                 ! EQ. 7
  1        0         1        0                              ! EQ. 8
 20 10 10 10 20 2         DELAYS
KNOWLEDGE                             5 35
MOTIVATION                            .7 1.3
LEARNING RATE                         0      .0075
TEACHING TIME                         0 .075
CONTROL TIME                          0 .075

. END OF DATA
```

```
CLASSROOM DYNAMICS MODEL

13-Nov-76       08:52



STUDENT INFORMATION

STUDENT     KNOWLEDGE      RACE     SEX        IQ      MOTIVATION
 ANDY          11.        WHITE    MALE       100.        1.00
 BEN            8.        BLACK    MALE       120.        1.20
 CINDY         12.        WHITE   FEMALE       97.        1.07
 DIANE          7.        BLACK   FEMALE      103.        0.90
 EGOR          10.        WHITE    MALE        90.        1.02
 FRAN           7.        WHITE   FEMALE       95.        0.78
 GREG          13.        BLACK    MALE        85.        0.93
 HENRY         10.        WHITE    MALE       100.        1.03
 IONA          10.        WHITE   FEMALE      100.        1.00
 JASON         10.        BLACK    MALE       101.        0.80
 KELLI         10.        BLACK   FEMALE      107.        0.86
 LARRY          9.        WHITE    MALE        93.        1.10
 MARY           9.        WHITE   FEMALE      100.        1.05
 NICK           9.        BLACK    MALE        92.        0.90
 OLIVE          9.        BLACK   FEMALE      100.        0.95
 PETER         11.        WHITE    MALE       105.        1.12
 QUINT         12.        BLACK    MALE       110.        1.04
 RITA          10.        WHITE   FEMALE      115.        1.17
 SANDY          9.        WHITE   FEMALE       78.        0.84
 TERRY         11.        BLACK   FEMALE      108.        1.14
 UXTON          8.        WHITE    MALE        86.        0.98
```

page 1 of sample output

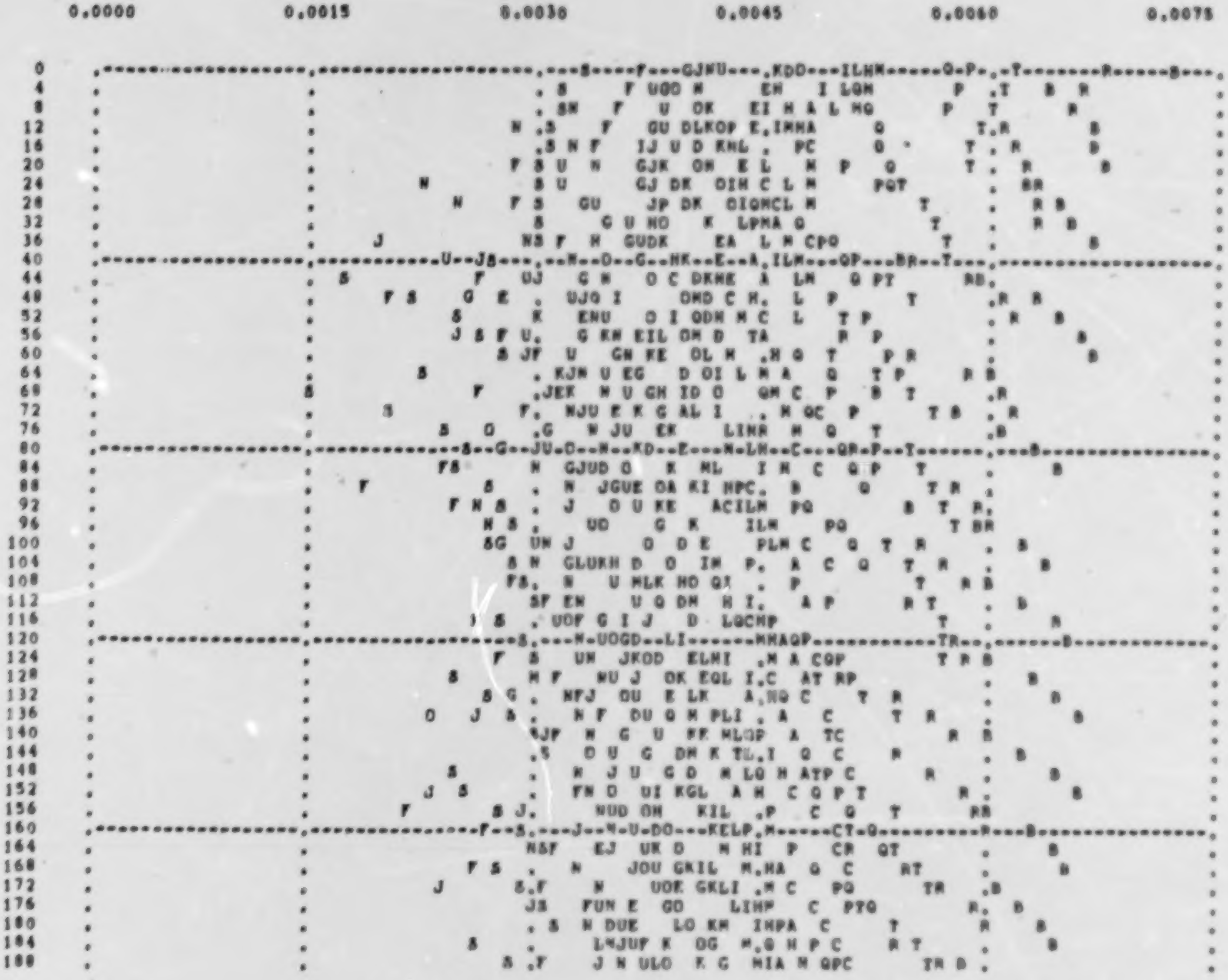
page 2 of sample output

```
WEIGHTS

    0.200000   -1.800000   -0.035000
    0.100000   -1.500000
    0.050000    0.075000    0.002500    0.001500    0.005000
    0.000055    0.000000    0.100000   10.000000
    1.000000    0.250000    1.000000    0.150000    0.300000   60.000000
    1.000000
    0.000000    0.000000
    1.000000    0.000000    1.000000    0.000000

DELAYS

  20.0  10.0  10.0  10.0  20.0   2.0
```

```
KNOWLEDGE                 13-Nov-76   08:52

             5.0000              11.0000             17.0000             23.0000             29.0000             35.0000

          0     ,----F--U--S---R--T--Q---G--------------,-------------------,-------------------,-------------------,
          4     .    F  UB  S  KR ,T  Q  G              .                   .                   .                   .
          8     .    F   U  SM  R ,AT  Q  G             .                   .                   .                   .
         12     .     F  UB SO  KR, PT  Q  G            .                   .                   .                   .
         16     .     F   UB SM  KR  PT CQ G            .                   .                   .                   .
         20     .     FD  U BSO  KHR APT CQ G           .                   .                   .                   .
         24     .     FD  U BNSM  K R AT  CQ G          .                   .                   .                   .
         28     .      F   U BSOM JK R PT  Q G          .                   .                   .                   .
         32     .      FD  U  SOM ,K  R PT  Q G         .                   .                   .                   .
         36     .      FD   U NS M,JK  RAPT  Q G        .                   .                   .                   .
         40     ,--- ----FD--U--SOM,JKI--RAPT-CQG-------,-------------------,-------------------,-------------------,
         44     .       FD  U  SBOM JK   RAPT CQG       .                   .                   .                   .
         48     .       F D  U SNO,MJKI   RP T QG       .                   .                   .                   .
         52     .       F D  U  S OM JKH   RP T QG      .                   .                   .                   .
         56     .        F D U  SNOBMJKI    RP TCQG     .                   .                   .                   .
         60     .        F D  U  S,OLMJKI    R PTCQG    .                   .                   .                   .
         64     .        F D  U  SN OMJKHI    R PTCQ    .                   .                   .                   .
         68     .         F D  U SN O MJKI     RP TCQ   .                   .                   .                   .
         72     .         F D  U S,N OLM KI      ARPTCQ,                   .                   .                   .
         76     .         F  D  U SN O LM KI      R PTCGQ                   .                   .                   .
         80     ,----------F-D--U-S-NO--MBKEI-----R-P-T-Q-------------------,-------------------,-------------------,
         84     .          F  D U ,SN O  MBKIH     R P.TCQ                  .                   .                   .
         88     .          F  D  U,S N D LMBKI     A RP TCQ                 .                   .                   .
         92     .          F   D  US N O  LMKEI     A RP GTQ                .                   .                   .
         96     .           F  D  U S N O  M KEI     A, R GT Q              .                   .                   .
        100     .           F    D U S N  O JMKE I     A  RG T Q            .                   .                   .
        104     .            F  D ,U S N O  LMKEHI      , A RG CTQ          .                   .                   .
        108     .            F   D, US N  O  LMKEHI     ,  A PR CTQ         .                   .                   .
        112     .             F D, U S N  O JM K BI     ,   A  R C T        .                   .                   .
        116     .             F  D U S N  O  JMKE IH    ,    A  R C T       .                   .                   .
        120     ,---------------F---D--U-S-N--O--LMKE-IH,-----A--PR-C-T-----,-------------------,-------------------,
        124     .               F  ,D  US N   O  LMKE IB      A  P RC QT    .                   .                   .
        128     .               F  , D US  N  O  JM KE I B     A  GR C T    .                   .                   .
        132     .                F , D  US  N  O  JM KE,IH B    A G PR C T  .                   .                   .
        136     .                F ,  D  US N  O  J M KE I  B      A GP RC QT                   .                   .
        140     .                 F, D  US  N  O  J MK,E IH B      A G P RCQT                   .                   .
        144     .                 F,   D  US N   O  J MK E IH  B    A G P RC T                  .                   .
        148     .                  F    D US  N   O  J MKE  IH   B    AG   P,RCQT               .                   .
        152     .                  F    D  U   N  O  J ,MKE  IH    B   AG    P  R QT            .                   .
        156     .                  F     D  U  N   O J , LM E IH     B   A G,P   R QT           .                   .
        160     ,-------------------,F-----D-US--N---O-J,--LM-E-IH-----B--AG,--P---R-T----------,-------------------,
        164     .                   .F      D  U   N   O J   MKE  IH       B  AG  P   R  T      .                   .
        168     .                   . F      D  U  N    O,J   LME  IH      B,AG   P   R QT      .                   .
        172     .                   . F      D SU  N    OJ    LME   I       ,BAG    P  CR QT    .                   .
        176     .                   .  F      D  U   N   O J   LME   IH     . BG      P  CR QT  .                   .
        180     .                   .  F       D SU  N  ,O J    LMK   IH    .   BG     P   C RQT,                   .
        184     .                   .   F      D  SU  N , O J   L ME   IH   .     G      P   CRQ,T                  .
        188     .                   .   F       D SU  N . O J    LMKE   IH  .      AG     P    C,RQ T               .
```

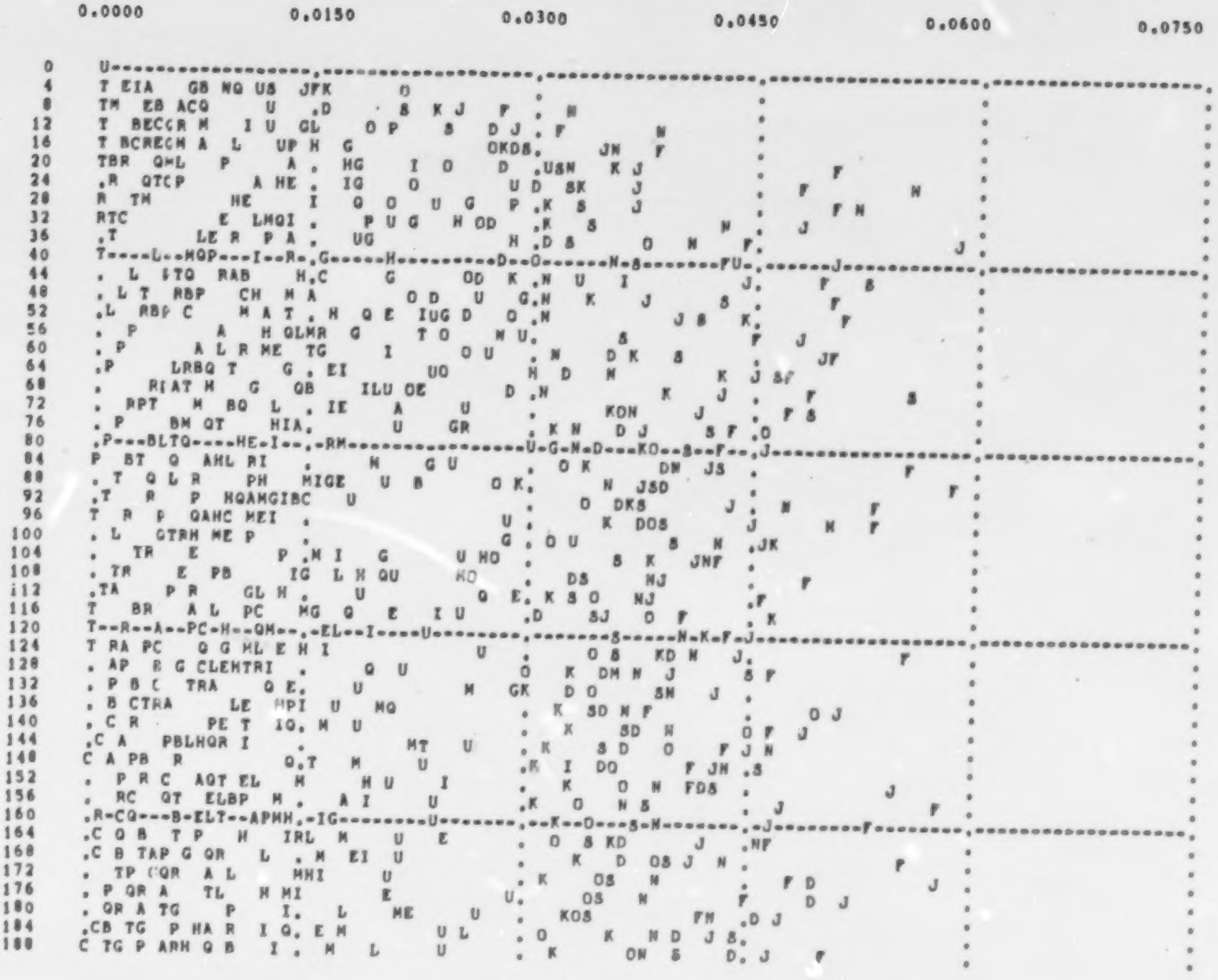
```
MOTIVATION              13-Nov-76   08:52

              0.7000             0.8200             0.9400             1.0600             1.1800              1.3000

  0     ,-----------F---J---,--S---K-----N----G-,-O----U--I--E-H-GM-,-C----L--P--T----R-,--B----------------,
  4     .           F     J .   S K      N     G, O    U  AI  EH QM . C    L   P  T    R,  B                .
  8     .           F     J .   S K      N     G O     U   I   EHQM , C    L   P T      R  B                .
 12     .           F      J.    S   K   DN    GO       U IA    EHQM, C    L    P  T    R  B                .
 16     .           F      J,   S    K   DN    G O      U  I A E HQM,   C   L   P T     .R B                .
 20     .           F       J   S     K N      .G O    U   I  A EH QM   C   L    P T    .R  B               .
 24     .           F       J S       K ND     . O G U     I  A EH Q,    C L     PT     .R  B               .
 28     .           F       .J  S     K ND     . O  GU     I  A E HQ,     C L    P T    . R B               .
 32     .           F       .J  S     K  ND    .  O  UG    I  A  EHQ,     C L    P  T   .  R B              .
 36     .            F      .J    S    K N D    . O  U   G I   A EHQM       C L    P T  .    R B            .
 40     ,------------F------,J----S----K-N-D----,-O--U----GI----AE-QM------C--L----P-T--,---R--B------------,
 44     .            F      .J    S      KND    . O  U     I     E QM     CL       P  T .   R B             .
 48     .            F      .J    S       ND    .O   U     I    EAHQM      LC     P    T    R  B            .
 52     .            F      .J    S      NKD    . O  U     IG    EHQM         L  CP    T    R   B           .
 56     .               F   .J    S      NKD    .O   U     I  G   HAQ      L    P      T     R  B           .
 60     .               F   .  J  S       NKD   .O  U      I   G EH O      L     P      .T   RB             .
 64     .                 F .  J  S       NK    .O   U      I   GEHM,Q      L    P C    .T   B R            .
 68     .                 F .  J  S       NK    O     U     I   GEHM,Q      L    P C    . T  BR             .
 72     .                  F,   J S      N  K   .O     U     I   GEM,QA      L   P  C   . T R  B            .
 76     .                  F,    JS     ND  K   .O       U    I  G M,QA      L P      C .T R   B            .
 80     ,-------------------F---S-----ND--K-----,O--------U----I--GM,-QA-----L---P-----C,-TR----B-----------,
 84     .                   F   S     N  D  K   O         U    I EMH    QA  L   P      C, TR    B           .
 88     .                   F    J S  N  D   K  O          U   I EMH     Q  L   P      C,  T    B           .
 92     .                   F    J   S   D   K  O          U     I MHG   QL     P       C R T   B           .
 96     .                   .F  J    NS  D   K .O           U    I MH    AQ     P       .R  T   B           .
100     .                   . F  J   N S D   K .O           U    I MEH    LQ    P       R C  T   B          .
104     .                   . F   J  NS  D   K .O            U   I M H   LA Q   P       R C T    B          .
108     .                   .   F  J N  SD     KO           U     I M HE LA Q    P      R C  T    B         .
112     .                   .    F J N  DS     KO           U     I M HG L  Q     P     R  C T    B         .
116     .                   .    F JN   DS     KO           U      I,MEH L  AQ    P     .R  C T    B        .
120     ,-------------------,-----F-N----D-S--KO,-----------U------I,MEH-GL--Q-----P----R---C-T----B--------,
124     .                   .      FN    D  S  O,            U     I,MEH L   Q     P    R    CT     B       .
128     .                   .      FN      DS OK,           U       IM EHLG   AQ  P     .R    CT     B      .
132     .                   .       NJ     DS OK,           U       ,M E LG   AQ P      .  R T       B      .
136     .                   .        N    D S OK,           U       MI  EL G  A Q       .   R CT      B     .
140     .                   .          N  D S OK,          U        MI  LH  G APQ       .    RCT      B     .
144     .                   .         NF  D S  O,           U       MI  LEH   GA P Q    .     RT        B   .
148     .                   .         NF D  S OK           U        MI   LEH   A PQ     .     R T       B   .
152     .                   .        J N D  S OK          U         MI   LH    A PQ     .      RCT    B     .
156     .                   .          JN D  SOK          U         MI   LH    A  PQ    .         RT  B     .
160     ,-------------------,-----------J-ND--S-,K---------U--------M-I--E------A--P-GQ-,---------RT---B----,
164     .                   .           JFN D SO, K         U       MI  ELH    A   P   Q,         RTC  B    .
168     .                   .            J ND SO,  K       U        M I   L H    A  P   GQ          RT B    .
172     .                   .            JF N SO    K     U         M I  LEH      A P    GQ          T B    .
176     .                   .             JFN   SO  K       U       .MI  L H      A  P  .GQ         TRB     .
180     .                   .             J FN  .SO      K U        .MI   L EH     A  P .G Q         TR     .
184     .                   .             J  N S  O      KU         MI    L EH      A  P,  Q         TR     .
188     .                   .              J FN .S  O      U        M,I    L EH        A  P  GQ      BTCR   .
```

LEARNING RATE 13-Nov-76 08:52



```
TEACHING TIME          13-Nov-76   08:52

      0.0000             0.0150             0.0300             0.0450             0.0600             0.0750

  0     .------------------.------------------.------------------.--U---------------.------------------.
  4     .                  .                  .            GCTRS QKU      D         .                  .
  8     .                  .                  .      G T RQSKIFBU, O  N             .                  .
 12     .                  .                  G  TQCJ SR MBU OP D,     N            .                  .
 16     .                  .             G    TC  J MS M PUL   O ,N                 .                  .
 20     .                  .         G     T  .JR SMKBLHI O DN  U,                  .                  .
 24     .                  .      G    O CTAJR,PSM  LIBO    D FU .         N        .                  .
 28     .                  .        GCT ARJ QESMHL O     PU   F  .       N          .                  .
 32     .                  .     GC TA RQ  EKS, BM PO HDU  F     ,N                 .                  .
 36     .                  . G C T Q  A ERISK P L    U HF O J   N,                  .                  .
 40     .------------------.G---TQ--A--E-IKPRMS.----BM-DOJ---N-----.U-----------------.------------------.
 44     .                  .  Q TA  E PR  ML  .HB   OD IN   S  U   .                  .                  .
 48     .                  . T   CARGP LQ HKM O I   NS  U     F    .                  .                  .
 52     .                  .   C GTP Q L   MJ I  DOS   U     F     .                  .                  .
 56     .                  .  GA PQ C   R TH MLO S  J     U        .                  .                  .
 60     .                  . G  Q C     T  EH IMKO SJ     D U F    .                  .                  .
 64     .                  G QP C    T       IMBO,J  KNH SU  F     .                  .                  .
 68     .                G . C QTR       MIO  JNHL  U         F S  .                  .                  .
 72     .              G   C QT R       A M  E J,L   NO  UD  FS    .                  .                  .
 76     .                 Q,PC  TG  A  M  EJRL , N    U  SF  O     .                  .                  .
 80     .--------------Q-PC-T---A----IGRMB-LM-.N-J----S-OF-U------.------------------.------------------.
 84     .             Q  P T    A  GIR MB     KLM      O SU  D    F                  .                  .
 88     .            Q  T . C PG R A I H  LM  .  ON SU     D       .   F              .                  .
 92     .               TQ . P RC A  IH   LM   KJ   S U    ND      .  F               .                  .
 96     .             TQ   .P R  A    I   EM  K,J    S   O  UDN    .  F               .                  .
100     .            Q  T  C   R    I  GEM B   , J   O S  N    U   F                  .                  .
104     .          Q  T C   . R   P G  EI B  H  JH OKLS    N U  F  .                  .                  .
108     .            T     RA Q G C  I      B J ,KO MS   N U       .F                 .                  .
112     .          T     A.PR    C Q    BH JK LM  ESON  UD         .F                 .                  .
116     .        T         P. APGQ      B H JI M E    S N OK  U   F .                 .                  .
120     .-------T--------R-.OG--------BH--I--ME,J-L---SNO----UD----F------------------.------------------.
124     .      T        R Q PA              I H M .JL   K S        DU .      F        .                  .
128     .            T PGC. R  Q     BIH E  JL. K  ON      MDUS    .  F               .                  .
132     .           T PC  R.QA      B HG  I E K.L     ON   U       . F                .                  .
136     .         T  C  R  A P   QBG   I  E  L K  M N   UJ  D   O F.                  .                  .
140     .          C TR   APQ   G    BI  E   LK  M   N J SU  DO    .   F              .                  .
144     .        C       GR .  GT    B I  E   L .     MJ   SN ODU  .   F              .                  .
148     .       C        RA TQ       B H       LE .   MJ     O   U   S.  F            .                  .
152     .          CRP T G A    B          EH LKM        N  OU   J SD.  F             .                  .
156     .        CR T   GP .      A B   E L M.      N O  J U SD    .           F      .                  .
160     .-----CR-Q---T----P----A-G------EHILM-.--------O-J---U-S-D.-------------F----.------------------.
164     .     C Q  T    PR . ACB         HI      H  E   OJ     US  D .         F      .                  .
168     .    C   T Q  R     AG B         H   I  KME     J   O   U   S.               F.                  .
172     .      T Q RP     G. A B          IH KME.        O   NU  S  J    D       F    .                  .
176     .     CQ RPT       A G          I   ML K E       O   N   SU .      D     F    .                  .
180     .    Q R T      P A GB          I          K M  E  O       JN SU.     D     F .                  .
184     .   C  T   RP Q  G ,A          I             M K    OL   J  N U .  S D      F .                  .
188     .  C T   RPQ    G A. B       HI          MEK    L    O   N  U . S    D       F.                  .
```

CONTROL TIME 13-Nov-76 08:52



```
DAY   ------------------------------------------- (SCALED) KNOWLEDGE--------------------------------------------- 13-Nov-76   08:52
      ANDY   BEN  CINDY DIANE  EGOR  FRAN  GREG HENRY  IONA JASON KELLI LARRY  MARY  NICK OLIVE PETER QUINT  RITA SANDY TERRY UXION
  0    110    80   120    70   100    70   130   100   100   100   100    90    90    90   ·90   110   120   100    90   110    80
  4    112    81   122    71   101    70   131   101   101   101   101    91    91    91    90   112   121   102    91   112    80
  8    114    83   124    72   103    71   133   103   103   102   103    93    93    92    92   115   124   105    92   115    82
 12    116    86   126    73   105    72   135   105   105   104   104    94    94    92    94   115   127   107    93   117    83
 16    118    88   129    74   107    73   137   106   106   106   106    96    96    94    95   118   130   110    94   120    84
 20    120    91   131    75   108    74   139   108   108   107   107    98    98    95    97   120   132   112    95   123    85
 24    122    92   133    76   110    75   141   110   110   109   109   100   100    95    98   123   135   115    96   125    86
 28    124    95   136    78   112    76   143   112   112   110   111   102   102    96   100   124   137   118    97   128    87
 32    127    98   138    79   114    77   145   113   114   111   113   103   104    98   101   127   140   121    99   131    89
 36    128   100   141    80   116    78   147   115   116   111   115   105   106    99   102   130   143   124   100   134    90
 40    131   102   142    81   118    79   150   117   118   113   117   107   108   100   104   133   146   126   101   137    90
 44    133   105   145    83   120    79   151   119   118   115   119   109   110   102   105   136   149   129   101   140    92
 48    135   107   148    84   120    79   152   121   120   116   120   111   111   103   107   138   149   133   102   143    93
 52    138   110   150    86   122    81   154   123   123   118   121   113   113   105   109   141   152   136   103   146    94
 56    141   114   152    87   124    82   157   126   124   119   123   115   115   106   111   144   155   138   104   147    95
 60    143   117   155    88   126    83   159   128   127   120   125   117   117   108   112   147   158   142   106   151    97
 64    146   119   158    90   128    84   162   129   129   122   126   119   119   109   114   150   162   145   106   154    98
 68    149   121   162    91   129    85   164   131   131   124   128   120   122   111   116   153   164   149   106   158   100
 72    150   124   165    93   131    86   167   134   133   125   130   122   124   112   117   156   168   153   107   162   101
 76    153   127   167    94   133    87   168   136   136   127   132   125   127   114   118   160   171   154   109   165   103
 80    156   131   171    95   136    88   170   139   138   128   134   127   128   115   120   163   175   158   110   169   103
 84    159   134   174    97   137    88   172   142   141   130   136   129   130   116   122   167   179   162   111   172   105
 88    161   135   177    98   139    89   175   144   143   132   139   132   133   118   124   168   183   166   112   177   107
 92    164   139   180    99   142    90   178 [illegible] 146 134   141   134   135   119   126   172   186   170   114   181   108
 96    167   143   184   101   144    91   181   149   148   136   143   137   138   120   127   176   190   174   115   185   110
100    171   147   187   103   147    93   182   152   151   137   144   139   141   122   130   179   194   178   116   188   111
104    174   151   191   104   149    94   185   153   153   140   146   140   143   123   132   181   198   182   118   193   112
108    177   154   193   106   152    95   188   156   156   142   149   142   144   125   134   185   200   187   119   197   114
112    181   158   196   108   152    96   191   159   159   144   152   145   147   127   136   189   202   190   121   202   116
116    184   163   201   109   155    97   193   162   161   146   152   148   150   128   137   192   207   195   122   206   117
120    187   167   205   111   158    99   196   165   164   148   154   150   153   130   139   197   211   199   123   211   119
124    191   171   209   113   161   100   199   168   167   150   157   153   156   132   142   201   216   204   125   216   120
128    195   175   212   115   163   101   203   171   170   153   160   155   156   134   144   205   218   207   126   219   122
132    198   180   217   117   166   103   204   174   172   154   163   158   159   135   146   209   223   212   127   224   124
136    202   185   221   118   168   104   207   177   176   155   165   160   162   137   146   212   225   217   129   229   126
140    206   189   226   120   171   105   210   180   179   157   168   163   165   139   148   216   230   222   131   233   128
144    210   194   231   122   174   107   214   184   183   160   171   166   167   141   151   221   236   226   132   236   130
148    214   199   235   124   177   109   218   187   183   162   174   169   170   143   153   225   239   232   133   241   132
152    217   205   239   125   180   110   221   189   187   162   176   172   174   145   155   230   245   238   135   247   134
156    219   209   244   127   183   110   225   192   191   165   179   175   177   147   158   234   250   244   136   253   135
160    224   215   249   129   186   112   227   196   194   168   183   178   180   149   161   237   256   250   138   257   137
164    228   221   254   132   188   113   231   200   198   170   185   180   183   150   163   242   262   253   140   263   140
168    233   227   260   134   191   114   235   204   201   173   188   184   186   152   165   248   266   259   141   269   142
172    236   231   264   134   195   116   239   207   205   173   192   187   190   155   168   253   272   266   143   276   144
176    241   237   269   136   197   118   243   211   208   176   195   191   194   157   171   258   278   272   145   281   146
180    246   244   275   138   200   120   247   215   212   179   198   193   196   159   174   262   284   279   147   287   148
184    249   250   281   141   204   122   251   219   216   182   201   195   200   161   177   268   288   284   148   294   150
188    254   255   287   143   207   123   256   223   220   184   205   199   204   164   179   274   294   292   150   301   153
```

page 5A-21

(SCALED) MOTIVATION — 13-NOV-76 08:52

DAY	ANDY	BEN	CINDY	DIANE	EGOR	FRAN	GREG	HENRY	IONA	JASON	KELLI	LARRY	MARY	NICK	OLIVE	PETER	QUINT	RITA	SANDY	TERRY	UXTON
0	100	120	106	90	102	78	93	103	100	80	86	109	105	90	95	112	104	117	84	114	98
4	99	120	107	90	102	78	93	102	100	80	86	110	104	90	94	112	104	117	84	114	98
8	100	119	107	90	102	77	93	103	100	81	86	110	105	90	94	112	104	118	84	114	98
12	101	119	107	90	103	77	94	104	100	81	87	110	105	90	94	112	104	118	84	114	98
16	101	119	108	89	103	77	94	104	100	81	87	110	105	90	95	113	104	118	84	114	98
20	102	120	108	90	103	77	94	104	100	81	88	111	105	89	95	114	105	118	84	114	98
24	102	120	108	90	103	77	96	104	100	82	88	110	105	89	95	114	105	118	83	115	97
28	102	120	109	90	103	77	97	104	100	82	87	111	105	89	95	114	105	119	81	115	97
32	102	121	109	91	104	78	98	104	100	82	87	111	105	90	95	114	105	119	81	116	97
36	102	121	110	90	103	78	99	104	100	82	88	111	106	89	95	114	105	120	84	116	97
40	103	122	110	90	103	78	99	105	100	82	88	112	105	89	95	114	105	120	85	116	97
44	103	122	110	90	103	78	100	105	100	82	89	111	105	89	95	114	105	121	85	116	97
48	103	122	112	90	103	78	100	105	100	82	89	111	105	89	94	114	105	120	85	117	97
52	105	123	113	90	104	78	101	104	100	82	89	111	105	89	95	114	105	121	85	118	97
56	105	123	114	90	104	79	102	104	100	82	89	111	106	89	94	114	106	121	85	118	97
60	106	121	114	90	103	79	102	104	100	82	90	111	105	89	94	114	105	121	85	118	97
64	106	121	114	90	104	80	103	104	100	83	90	112	105	89	94	114	106	121	85	118	97
68	106	121	115	89	104	80	103	104	101	83	90	112	105	89	94	114	106	121	85	119	97
72	106	121	116	89	104	80	104	105	102	84	90	112	105	89	94	114	106	120	85	118	98
76	107	122	116	89	104	81	104	105	102	85	91	112	105	88	94	113	106	119	85	118	99
80	107	122	117	89	104	81	104	105	103	85	90	111	105	88	94	113	107	120	84	119	100
84	108	122	117	89	105	81	105	105	103	85	91	110	105	88	94	112	108	119	84	119	99
88	109	122	117	89	105	81	106	105	103	85	92	110	105	87	94	112	108	119	86	119	100
92	109	122	118	90	105	82	106	106	104	84	92	109	105	87	94	113	109	119	87	120	100
96	109	123	118	89	106	82	106	106	104	84	92	109	105	87	94	112	109	118	88	120	100
100	109	123	119	89	106	83	107	107	104	84	92	109	105	87	94	113	110	118	88	120	101
104	109	123	118	89	107	83	107	107	104	85	92	109	105	87	94	113	110	117	88	120	101
108	109	123	118	89	107	84	107	107	105	85	93	109	105	87	94	113	110	117	89	120	101
112	110	124	119	89	107	84	107	107	105	86	93	109	106	87	93	114	110	118	89	121	101
116	110	124	120	89	107	85	107	107	105	86	93	109	106	87	93	114	111	118	89	121	101
120	111	124	120	89	107	85	108	108	105	86	93	109	106	86	93	114	111	118	90	121	101
124	111	125	121	89	107	86	109	108	105	86	93	109	106	86	93	114	112	118	91	121	101
128	112	125	120	90	107	86	109	108	106	87	93	108	106	86	92	114	112	118	91	121	101
132	111	125	121	89	107	86	109	108	106	87	93	108	106	86	92	113	112	119	90	121	101
136	112	126	121	89	108	87	110	108	106	87	93	108	106	87	92	113	113	120	90	121	100
140	112	126	121	89	108	87	110	109	106	87	93	108	106	87	93	113	114	121	91	121	100
144	112	126	122	90	109	88	112	109	106	88	93	108	105	87	93	113	114	121	91	122	100
148	112	126	122	90	109	88	113	109	106	88	93	109	105	88	93	113	114	121	92	122	100
152	112	126	122	90	109	88	114	109	106	88	93	109	106	88	93	114	115	122	92	123	100
156	112	126	123	90	109	88	115	109	106	88	93	108	105	88	93	114	115	123	92	123	99
160	113	126	124	90	109	89	116	109	107	89	94	109	105	90	93	115	116	123	92	124	100
164	113	126	124	91	109	89	117	109	106	89	95	109	106	90	93	115	117	123	92	123	100
168	114	126	124	91	109	89	117	110	107	89	95	109	106	91	93	115	117	123	92	124	100
172	114	126	124	91	110	90	118	110	106	89	96	109	106	91	93	115	118	124	93	124	100
176	114	125	124	91	110	91	118	110	106	90	97	109	106	91	94	116	118	125	94	124	100
180	114	125	125	91	110	91	118	111	107	90	98	109	106	92	95	117	119	125	94	125	100
184	114	125	125	91	110	92	119	111	106	90	98	109	105	92	95	117	119	125	94	125	99
188	116	124	125	92	111	92	119	111	106	90	99	109	105	92	96	118	120	126	94	125	100

```
DAY    ----------- --------------------------------- (SCALED) LEARNING RATE------------------------------------- 13-Nov-76   08:52
      ANDY  BEN   CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXTON
   0    50    72    51    46    45    37    39    51    50    40    46    51    52    41    47    58    57    67    32    61    42
   4    48    64    50    45    45    35    38    45    48    39    40    50    51    40    38    57    51    66    31    60    37
   8    48    65    46    44    44    35    37    46    44    38    41    49    50    32    40    57    52    65    31    59    37
  12    47    66    47    39    43    34    37    47    45    37    41    40    46    28    41    43    52    60    30    59    38
  16    47    66    47    40    43    34    37    42    37    37    41    43    47    31    38    47    52    61    30    58    38
  20    43    67    48    41    43    28    36    41    40    37    38    45    47    34    40    50    53    62    30    58    31
  24    44    62    44    38    39    30    36    43    42    37    39    46    48    22    41    52    53    62    30    53    31
  28    45    64    46    40    40    28    33    45    43    37    40    46    48    25    42    38    44    63    29    55    33
  32    46    65    46    41    41    30    34    37    44    30    41    43    44    29    38    44    47    63    29    54    35
  36    42    66    47    37    42    31    35    33    45    19    38    45    46    29    29    48    49    57    29    56    36
  40    44    54    38    39    42    32    36    38    45    25    39    46    47    32    34    51    50    55    26    57    24
  44    45    59    38    40    42    26    33    42    29    30    40    46    47    34    37    52    51    58    17    53    29
  48    41    62    42    41    27    20    25    44    35    32    33    47    40    32    39    49    33    60    22    54    32
  52    43    64    45    41    32    24    30    42    39    34    30    47    43    34    37    51    40    61    24    50    34
  56    44    66    38    41    36    27    33    44    37    25    34    38    40    35    39    52    44    50    26    44    29
  60    45    66    42    38    38    29    35    45    40    29    37    41    43    35    40    53    47    55    27    49    32
  64    46    60    45    39    35    31    36    33    42    32    31    43    44    32    41    53    49    58    22    52    34
  68    47    52    47    40    31    26    37    38    39    31    32    36    46    34    41    49    45    60    14    54    36
  72    40    57    48    36    35    29    38    41    41    33    36    40    46    32    34    51    47    61    20    56    33
  76    42    61    45    35    37    31    31    43    43    34    38    42    47    33    26    52    49    44    23    52    35
  80    44    63    47    37    39    32    27    45    44    30    36    44    42    34    32    52    50    51    25    54    30
  84    46    64    48    34    33    23    32    46    45    32    39    41    40    29    36    52    51    55    25    55    33
  88    38    47    44    34    36    19    35    42    41    34    40    43    43    31    38    43    51    57    26    56    35
  92    42    54    43    36    38    24    36    44    43    32    38    44    44    26    35    47    48    59    27    57    37
  96    44    59    46    38    40    28    38    45    44    34    40    45    45    27    34    49    50    60    28    57    33
 100    46    62    48    39    41    30    27    46    45    32    29    45    46    30    37    45    51    55    27    52    29
 104    47    64    49    37    41    32    32    35    41    34    34    33    42    29    39    44    51    57    28    54    33
 108    47    59    40    38    42    28    35    39    43    35    38    37    36    31    40    47    42    58    29    55    35
 112    47    62    40    39    32    30    36    42    44    36    40    40    40    33    36    49    38    54    29    56    36
 116    44    64    44    40    36    32    34    44    36    37    26    42    43    31    32    45    43    56    27    56    31
 120    46    65    47    36    38    34    36    45    40    34    32    39    44    32    35    48    47    57    29    57    34
 124    47    60    48    38    40    28    37    42    42    35    36    41    45    33    37    49    49    58    30    57    32
 128    47    62    46    39    41    31    38    44    44    36    39    42    30    34    39    50    42    49    24    48    35
 132    44    64    48    40    38    33    28    45    40    33    41    40    36    32    35    51    46    53    27    51    36
 136    46    65    49    36    40    34    33    41    43    25    39    42    40    33    23    41    38    56    28    53    37
 140    47    59    50    38    41    31    35    44    44    30    40    43    43    34    29    45    44    57    29    49    38
 144    47    62    50    39    42    33    37    45    45    33    41    43    40    31    33    47    48    54    30    43    35
 148    49    64    50    40    39    35    39    46    33    35    42    44    42    32    36    49    44    56    24    48    36
 152    43    6     47    32    41    32    40    38    38    23    39    41    44    34    35    50    48    58    25    52    37
 156    38    6     48    36    42    21    41    38    41    29    41    42    45    34    37    45    50    59    27    54    35
 160    42    6     49    38    42    27    35    41    43    32    42    43    46    35    38    44    52    60    29    51    36
 164    44    6     50    39    34    31    38    44    44    35    28    39    42    30    39    47    53    51    30    53    37
 168    46    6     51    40    37    26    40    45    40    36    40    41    44    32    37    49    48    54    27    55    38
 172    43    6     47    29    40    31    41    42    43    24    42    42    45    34    38    50    51    57    29    56    38
 176    45    6     48    30    37    33    38    44    44    29    39    43    45    35    39    50    52    58    30    52    36
 180    46    6     49    35    36    35    40    46    45    33    41    39    42    33    40    46    53    59    31    54    36
 184    44    6     50    37    39    37    41    47    45    35    38    34    44    34    41    48    46    54    26    55    37
 188    46    5     50    36    41    30    42    47    46    34    41    38    45    35    38    49    49    57    28    56    37
```

```
DAY     ------------------------------------------ (SCALED) TEACHING TIME------------------------------------------ 13-NOV-76   08:52
         ANDY   BEN   CINDY DIANE EGOR   FRAN   GREG  HENRY IONA  JASON KELLI LARRY MARY   NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXTON
   0      47     47     47     47    47     47     47     47    47     47    47    47    47     47    47    47    47    47    47    47    47
   4      41     46     41     44    42     44     40     45    42     42    45    43    43     43    50    41    45    42    43    41    46
   8      37     43     39     42    38     43     34     41    42     38    41    40    40     49    46    37    39    38    40    37    43
  12      33     40     34     44    35     40     29     37    37     35    37    42    39     49    41    42    34    37    37    33    40
  16      30     38     30     43    32     39     26     38    42     32    34    39    37     45    42    38    30    34    35    30    38
  20      30     35     27     40    30     44     23     37    38     30    35    37    34     41    39    34    26    31    33    27    44
  24      27     35     26     40    32     42     20     34    34     28    32    34    32     52    34    30    23    28    31    26    43
  28      25     33     23     38    29     43     22     31    31     26    30    32    30     50    33    38    28    26    33    24    40
  32      23     32     21     37    27     40     20     36    29     32    28    34    33     46    35    34    25    24    29    22    37
  36      23     29     18     37    25     38     17     37    26     41    28    31    30     43    40    30    22    26    27    20    35
  40      21     34     23     36    24     36     15     34    25     37    26    29    28     40    36    26    19    27    29    18    45
  44      20     31     23     34    22     40     17     30    35     33    24    27    26     36    33    24    17    25    39    19    41
  48      21     28     20     31    30     42     22     27    31     29    27    25    28     34    30    23    25    22    35    17    37
  52      20     27     18     32    28     41     20     29    29     28    33    25    27     33    32    21    22    20    34    21    36
  56      18     26     22     31    26     38     17     26    29     34    30    28    28     31    30    19    19    24    33    26    38
  60      18     27     20     37    26     40     16     26    28     33    29    29    29     31    30    18    15    23    32    24    38
  64      17     28     18     36    27     39     14     34    27     31    33    28    27     33    29    17    16    21    36    21    36
  68      16     30     18     34    29     40     13     31    26     30    32    32    25     30    27    18    18    19    43    19    34
  72      23     29     14     37    28     40     12     29    25     28    30    30    25     33    34    17    16    18    40    17    36
  76      21     27     16     38    26     38     19     26    23     27    28    28    24     31    40    15    14    27    37    18    34
  80      19     26     15     37    24     38     24     25    22     32    30    27    28     30    37    14    12    25    35    16    39
  84      18     25     13     39    29     45     21     23    21     30    28    28    29     34    34    13    11    22    37    15    37
  88      22     32     16     40    27     48     18     24    23     28    26    27    27     32    32    18    10    20    34    13    35
  92      20     30     18     39    26     47     16     23    22     30    30    26    26     38    36    16    13    18    36    12    34
  96      19     29     17     40    26     47     15     23    22     30    29    26    27     41    37    15    12    17    34    11    39
 100      17     27     14     38    24     44     23     22    21     31    35    25    35     38    36    17    10    18    36    12    42
 104      16     25     13     39    23     42     20     30    23     27    32    33    27     38    32    19     9    16    34    11    39
 108      15     26     20     38    22     46     18     28    22     28    30    31    32     36    31    17    16    14    33    10    38
 112      14     25     20     38    32     45     16     26    22     27    28    30    30     25    34    15    21    16    33     8    37
 116      16     23     17     36    30     43     17     24    26     26    27    28    28     35    36    17    18    14    34     7    39
 120      15     23     15     41    29     44     16     24    26     30    35    32    28     36    36    15    15    13    35     6    40
 124      15     25     13     42    28     51     15     26    25     30    34    31    28     36    35    14    13    12    36     6    43
 128      13     23     14     40    26     48     13     24    23     28    31    29    38     34    33    12    18    16    40    11    40
 132      16     21     12     39    27     46     24     23    26     30    29    30    35     35    34    11    15    14    39     9    38
 136      15     20     10     39    25     44     21     24    24     37    2[illegible] 28 33   33    43    16    19    12    37     7    36
 140      14     23      9     41    26     48     18     24    24     36    29    28    32     35    42    14    18    11    38    10    38
 144      13     21      7     41    25     48     16     23    23     34    28    28    34     39    40    13    13    13    38    17    42
 148      12     20      6     41    28     48     15     22    33     33    27    27    32     38    38    11    14    11    43    14    41
 152      14     18      8     43    26     47     13     26    29     41    28    28    30     36    38    10    12    10    43    11    39
 156      20     21      7     42    25     55     12     27    27     39    27    27    28     35    36    13     9     8    41     9    40
 160      18     20      6     44    25     55     20     26    26     38    27    27    28     36    36    15     8     6    42    11    41
 164      16     18      5     43    31     53     17     24    25     35    29    30    30     41    35    12     6    13    41     9    40
 168      15     17      4     44    30     58     15     24    27     35    29    29    29     40    38    11     9    11    44     7    41
 172      16     18      6     49    28     56     14     26    25     45    27    28    28     39    36     9     7     9    42     6    40
 176      15     16      5     51    31     55     16     24    24     42    30    28    27     39    36     8     5     7    43     8    43
 180      13     15      4     51    33     55     15     24    24     41    29    31    31     42    36    11     4     5    43     7    44
 184      15     13      3     49    31     54     13     22    22     38    31    35    29     41    35     9    11     9    48     5    43
 188      14     16      2     51    30     59     12     22    22     40    30    34    29     41    38     8     9     7    47     4    43
```

page 5A-24

```
DAY   ----------------------------------- (SCALED) CONTROL TIME--------------------------------------- 13-NOV-76   08:52
       ANDY  BEN  CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXTO
  0       0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
  4       3    7     0     8     2    15     7     8     3    14    15     0     0     8    21     0     9     0    11     0    10
  8       5    4     6    15     3    27    12     7    12    24    22     0     1    32    20     0     7     0    21     0    12
 12       6    3     4    27     3    31    13     5    10    28    24    14     7    37    19    20     5     5    24     0    11
 16       7    2     2    28     4    38    16    14    27    34    27     9     5    34    26    13     4     3    29     0    12
 20      13    1     1    28     5    50    18    17    21    36    34     6     5    32    24     8     4     2    31     0    30
 24      11    5     5    30    13    47    17    13    17    36    33     4     4    55    21     5     3     1    32     4    28
 28       9    3     3    28    10    49    25    10    14    36    31     2     3    51    20    28    19     1    33     2    23
 32       9    2     2    27     8    42    21    24    13    47    30    11    11    42    26    18    12     0    33     1    20
 36      13    1     1    30     8    44    18    28    12    58    33     7     9    40    37    12    10     9    32     1    18
 40      11   15    15    27     6    43    16    19    11    50    30     4     6    35    30     8     7    13    36     0    43
 44       9   10    15    25     6    49    20    14    35    44    28     3     5    31    25     9     6     8    52     5    33
 48      14    6    10    23    26    50    29    11    26    37    36     2    13    30    21     7    26     5    42     3    25
 52      12    4     6    25    20    50    23    16    22    39    44     1    10    31    28     4    19     3    41    13    23
 56       9    2    15    23    14    44 [illegible] 11 22    47    36    14    15    27    23     3    13    15    36    22    28
 60       7    1    10    35    12    49    16    10    20    49    36     9    11    31    25     2    10    10    40    14    26
 64       6    7     6    32    16    47    13    29    17    45    42     6     9    35    23     1     8     6    46     9    23
 68       5   14     4    28    22    48    11    21    18    42    39    19     7    30    21     5    14     4    55     6    19
 72      20    9     2    35    17    46    10    16    16    41    35    12     7    37    36     3    10     2    48     4    25
 76      14    6     8    36    13    43    25    12    13    37    31     8     6    33    45     2     8    25    42     8    28
 80      10    4     5    34    11    42    31    10    12    45    37     5    17    32    38     1     6    18    40     5    30
 84       8    2     3    39    23    55    23     8    11    41    33     9    19    40    32     0     5    10    42     3    24
 88      16   27    11    38    17    58    16    12    15    37    29     6    14    35    27    11     4     7    37     2    28
 92      11   14    11    36    13    53    12     9    13    43    36     4    12    47    33     7    10     4    37     1    18
 96       8    9     9    37    12    52    10     9    13    45    35     3    11    50    37     4     7     3    39     1    28
100       5    6     6    33     9    46    28     7    11    45    46     2     9    42    31    11     5     7    40     5    33
104       4    4     4    35     7    42    20    26    17    41    38    25    15    42    27    13     3     4    35     3    25
108       3    5    20    33     6    49    14    19    13    38    33    17    25    38    26     8    20     3    33     2    21
112       2    6    18    32    28    45    10    13    11    38    31    11    18    37    34     5    26     7    33     1    18
116       7    4    12    30    20    48    15    11    23    35    46     8    14    37    37    10    18     4    34     1    25
120       5    5     8    40    16    43    12     9    19    45    42    16    12    40    36     7    11     3    36     0    23
124       3 [illegible] 5 39    13    55     9    14    16    44    38    12    11    41    34     4     7     2    35     0    27
128       2 [illegible] 8 34    10    46     6    10    12    39    32     8    36    37    30     3    19    12    44    11    21
132       8 [illegible] 5 32    13    42    29     8    18    42    29    12    26    39    34     2    12     7    38     7    18
136       6 [illegible] 3 34    11    38    20    13    17    50    32    10    20    36    49    14    21     5    34     4    17
140       4 [illegible] 2 37     9    46    14    10    13    49    32     8    16    40    44     9    13     3    36    11    18
144       2 [illegible] 1 36     7    43     9     8    11    45    31     7    22    46    39     6     9     8    35    23    26
148       1 [illegible] 1 35    14    41     6     7    33    42    30     6    18    43    36     3    14     6    45    15    23
152       8 [illegible] 6 41    11    41     4    19    24    54    33    11    15    38    36     2     9     4    42    10    20
156      10 [illegible] 4 38     9    58     2    19    19    47    31     9    13    37    33    11     6     2    37     6    24
160      12 [illegible] 2 38     8    52    17    14    16    46    32     8    13    38    34    13     4     1    37     9    23
164       8 [illegible] 1 36    24    46    11    11    13    42    36    15    17    45    32     8     2    14    34     6    21
168       5 [illegible] 1 36    18    55     7     9    19    41    34    12    16    43    38     5     9     9    39     4    21
172       9 [illegible] 5 49    14    48     4    15    16    57    31    10    14    39    35     3     5     6    35     2    20
176       6 [illegible] 3 49    20    44    10    12    14    51    35     9    14    38    34     2     3     4    35     8    29
180       4 [illegible] 2 45    22    42     6    10    14    47    32    17    21    42    33    10     2     2    34     5    27
184       8 [illegible] 1 40    16    38     4     8    13    42    35    25    18    38    31     7    14    10    44     3    24
188       6   10     1    44    13    50     2     7    13    46    32    20    17    37    37     4     9     7    40     2    24
```

```
DAY        CLASS    AVGKNO    AVGMOT    AVGLRN    AVGLIK          13-Nov-76   08152
  0        0.900     9.762     0.994     0.005     0.000
  4        0.900    10.119     0.992     0.005     0.033
  8        0.899    10.155     1.003     0.005     0.056
 12        0.898    10.423     1.006     0.004     0.075
 16        0.897    10.439     1.000     0.004     0.091
 20        0.896    10.749     1.005     0.004     0.105
 24        0.895    10.886     1.011     0.004     0.117
 28        0.894    10.753     1.003     0.004     0.126
 32        0.892    11.243     1.022     0.004     0.129
 36        0.891    11.332     1.012     0.004     0.134
 40        0.889    11.718     1.008     0.004     0.141
 44        0.888    11.868     1.039     0.004     0.147
 48        0.886    11.641     1.018     0.004     0.138
 52        0.884    12.013     1.010     0.004     0.143
 56        0.883    12.352     1.012     0.004     0.141
 60        0.881    12.555     1.021     0.004     0.148
 64        0.879    12.945     1.025     0.004     0.151
 68        0.878    12.952     1.025     0.004     0.151
 72        0.876    13.478     1.036     0.004     0.152
 76        0.874    13.122     1.016     0.004     0.153
 80        0.873    13.685     1.027     0.004     0.150
 84        0.871    14.367     1.046     0.004     0.160
 88        0.870    13.938     1.013     0.004     0.153
 92        0.868    14.570     1.046     0.004     0.155
 96        0.866    14.799     1.036     0.004     0.157
100        0.865    14.524     1.036     0.004     0.159
104        0.863    15.195     1.033     0.004     0.150
108        0.861    14.903     1.028     0.004     0.159
112        0.860    15.723     1.047     0.004     0.155
116        0.858    15.497     1.041     0.004     0.156
120        0.857    16.500     1.051     0.004     0.156
124        0.855    16.793     1.056     0.004     0.162
128        0.853    17.005     1.055     0.004     0.159
132        0.852    17.047     1.063     0.004     0.157
136        0.850    16.883     1.047     0.004     0.158
140        0.849    17.613     1.042     0.004     0.161
144        0.847    18.117     1.066     0.004     0.158
148        0.846    18.19      1.062     0.004     0.162
152        0.844    19.312     1.087     0.004     0.162
156        0.843    19.051     1.068     0.004     0.160
160        0.841    19.222     1.067     0.004     0.157
164        0.840    19.734     1.074     0.004     0.154
168        0.839    20.367     1.085     0.004     0.168
172        0.837    20.776     1.093     0.004     0.164
176        0.836    20.999     1.080     0.004     0.158
180        0.835    20.832     1.076     0.004     0.157
184        0.833    21.080     1.078     0.004     0.159
188        0.832    22.088     1.086     0.004     0.163

    DAYS ABSENT

      ANDY  BEN   CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINY RITA  SANDY TERRY UR
        14    16    17    18    19    33    21    22    19    28    24    18    17    25    26    18    21    16    22    14    2
```

```
LIKING MATRIX    DAY= 30        13-Nov-76      08:52


        ANDY  BEN   CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXT
ANDY       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
BEN        0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
CINDY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
DIANE      0     0     0     0     0     0     0     0     0     0     0     0     0     0   304     0     0     0     0     0
EGOR       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
FRAN       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   322     0
GREG       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   367     0     0     0
HENRY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
IONA       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
JASON      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
KELLI      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
LARRY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
MARY       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
NICK       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
OLIVE      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
PETER      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
QUINT      0     0     0     0     0     0   307     0     0     0     0     0     0     0     0     0     0     0     0     0
RITA       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
SANDY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
TERRY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
UXTON      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
```

```
LIKING MATRIX    DAY= 60        13-NOV-76       08:52


        ANDY   BEN CINDY DIANE  EGON  FRAN  GREG HENRY  IONA JASON KELLI LARRY  MARY  NICK OLIVE PETER QUINT  RITA SANDY TERRY   UXT
ANDY       0     0     0     0   306     0     0   320     0     0     0     0     0     0     0   350     0     0     0     0
BEN        0     0     0     0     0     0     0     0     0     0     0     0     0   306     0     0     0     0     0     0
CINDY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   347     0     0
DIANE      0     0     0     0     0   306     0     0     0     0     0     0     0     0   392     0     0     0     0     0
EGON       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
FRAN       0     0     0   338     0     0     0     0     0     0     0     0     0     0     0     0     0     0   402     0
GREG       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   472     0     0     0
HENRY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
IONA       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
JASON      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
KELLI      0     0     0     0     0     0     0     0     0     0     0     0     0     0   328     0     0     0     0     0
LARRY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
MARY       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
NICK       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
OLIVE      0     0     0     0     0     0     0     0     0     0   328     0     0     0     0     0     0     0     0     0
PETER    367     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
QUINT      0     0     0     0     0     0   436     0     0     0     0     0     0     0     0     0     0     0     0     0
RITA       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
SANDY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
TERRY      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
UXTON      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
```

```
LIKING MATRIX    DAY= 90        13-Nov-76       08:52



        ANDY  BEN   CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXT
ANDY         0     0     0     0     0     0     0   311     0     0     0     0     0     0     0   359     0     0     0     0
BEN          0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
CINDY        0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   431     0   326
DIANE        0     0     0     0     0     0     0     0     0     0     0     0     0     0   406     0     0     0     0     0
EGOR         0     0     0     0     0     0     0   313     0     0     0   315     0     0     0     0     0     0     0     0
FRAN         0     0     0   326     0     0     0     0     0     0     0     0     0     0     0     0     0     0   356     0
GREG         0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   470     0     0     0
HENRY        0     0     0     0   312     0     0     0     0     0     0   305     0     0     0     0     0     0     0     0
IONA         0     0     0     0     0     0     0     0     0     0     0     0   346     0     0     0     0     0     0     0
JASON        0     0     0     0     0     0     0     0     0     0     0     0     0   336     0     0     0     0     0     0
KELLI        0     0     0     0     0     0     0     0     0     0     0     0     0     0   325     0     0     0     0     0
LARRY        0     0     0     0   331     0     0   321     0     0     0     0     0     0     0     0     0     0     0     0
MARY         0     0     0     0     0     0     0     0   352     0     0     0     0     0     0     0     0     0     0     0
NICK         0     0     0     0     0     0     0     0     0   360     0     0     0     0     0     0     0     0     0     0
OLIVE        0     0     0     0     0     0     0     0     0     0   333     0     0     0     0     0     0     0     0     0
PETER      389     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
QUINT        0     0     0     0     0     0   490     0     0     0     0     0     0     0     0     0     0     0     0     0
RITA         0     0   369     0     0     0     0     0   319     0     0     0     0     0     0     0     0     0     0     0
SANDY        0     0     0     0     0     0     0     0     0     0     0     0   334     0     0     0     0     0     0     0
TERRY        0     0   340     0     0     0     0     0     0     0     0     0     0     0     0     0     0   346     0     0
UXTON        0     0     0     0     0     0     0     0     0     0     0   345     0     0     0     0     0     0     0     0
```

```
LIKING MATRIX    DAY=120        13-Nov-76      08:52



             ANDY  BEN   CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXT
      ANDY      0     0     0     0     0     0     0   339     0     0     0     0     0     0     0   386     0     0     0     0
      BEN       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
      CINDY     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   531     0   384
      DIANE     0     0     0     0     0   386     0     0     0     0     0     0     0     0   394     0     0     0   372     0
      EGOR      0     0     0     0     0     0     0   342     0     0     0   315     0     0     0     0     0     0     0     0
      FRAN      0     0     0   461     0     0     0     0     0     0     0     0     0     0     0     0     0     0   552     0
      GREG      0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   310   404     0     0     0
      HENRY   307     0     0     0   339     0     0     0     0     0     0   326     0     0     0     0     0     0     0     0
      IONA      0     0     0     0     0     0     0     0     0     0     0     0   364     0     0     0     0     0     0     0
      JASON     0     0     0     0     0     0     0     0     0     0     0     0     0   360     0     0     0     0     0     0
      KELLI     0     0     0     0     0     0     0     0     0     0     0     0     0     0   312     0     0     0     0     0
      LARRY     0     0     0     0   328     0     0   342     0     0     0     0     0     0     0     0     0     0     0     0
      MARY      0     0     0     0     0     0     0     0   369     0     0     0     0     0     0     0     0     0     0     0
      NICK      0     0     0     0     0     0     0     0     0   404     0     0     0     0     0     0     0     0     0     0
      OLIVE     0     0     0     0     0     0     0     0     0     0   322     0     0     0     0     0     0     0     0     0
      PETER   425     0     0     0     0     0     0   300     0     0     0     0     0     0     0     0     0     0     0     0
      QUINT     0     0     0     0     0     0   468     0     0     0     0     0     0     0     0   308     0     0     0     0
      RITA      0     0   477     0     0     0     0     0   304     0     0     0     0     0     0     0     0     0     0   337
      SANDY     0     0     0     0     0   359     0     0     0     0     0     0   341     0     0     0     0     0     0     0
      TERRY     0     0   419     0     0     0     0     0     0     0     0     0     0     0     0     0     0   434     0     0
      UXTON     0     0     0     0     0     0     0     0     0     0     0   308     0     0     0     0     0     0     0     0
```

page 5A-3U

```
LIKING MATRIX    DAY=150        13-NOV-76      08:52



          ANDY   BEN CINDY DIANE  EGOR  FRAN  GREG HENRY  IONA JASON KELLI LARRY  MARY  NICK OLIVE PETER QUINT  RITA SANDY TERRY   UXT
ANDY         0     0     0     0     0     0     0   364     0     0     0     0     0     0     0   635     0     0     0     0
BEN          0     0     0     0     0     0   342     0     0     0     0     0     0     0     0     0     0     0     0     0
CINDY        0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   600     0   364
DIANE        0     0     0     0     0   440     0     0     0     0     0     0     0     0   444     0     0     0   384     0
EGOR         0     0     0     0     0     0     0   387     0     0     0   377     0     0     0     0     0     0     0     0
FRAN         0     0     0   550     0     0     0     0     0     0     0     0     0     0     0     0     0     0   550     0     3
GREG       340   362     0     0     0     0     0     0     0     0     0     0     0     0     0   320   376     0     0     0
HENRY      330     0     0     0   382     0     0     0     0     0     0   349     0     0     0     0     0     0     0     0
IONA         0     0     0     0     0     0     0     0     0     0     0     0   344     0     0     0     0     0     0     0
JASON        0     0     0     0     0     0     0     0     0     0     0     0     0   337     0     0     0     0     0     0
KELLI        0     0     0     0     0     0     0     0     0     0     0     0     0     0   375     0     0     0     0     0
LARRY        0     0     0     0   391     0     0   366     0     0     0     0     0     0     0     0     0     0     0     0
MARY         0     0     0     0     0     0     0     0   347     0     0     0     0     0     0     0     0     0     0     0
NICK         0     0     0     0     0     0     0     0     0   387     0     0     0     0     0     0     0     0     0     0     3
OLIVE        0     0     0   324     0     0     0     0     0     0   387     0     0     0     0     0     0     0     0     0
PETER      485     0     0     0     0     0   312   315     0     0     0     0     0     0     0     0     0     0     0     0
QUINT      313     0     0     0     0     0   475     0     0     0     0     0     0     0     0   366     0     0     0     0
RITA         0     0   363     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   363
SANDY        0     0     0   311     0   365     0     0     0     0     0     0     0     0     0     0     0     0     0     0
TERRY        0     0   435     0     0     0     0     0     0     0     0     0     0     0     0     0     0   460     0     0
UXTON        0     0     0     0     0     0     0     0     0     0     0   357     0   362     0     0     0     0     0     0
```

```
LIKING MATRIX   DAY=180       13-NOV-76      08:52


         ANDY  BEN   CINDY DIANE EGOR  FRAN  GREG  HENRY IONA  JASON KELLI LARRY MARY  NICK  OLIVE PETER QUINT RITA  SANDY TERRY UXT
ANDY        0     0     0     0     0     0     0   353     0     0     0     0     0     0     0   410     0     0     0     0
BEN       323     0     0     0     0     0   451     0     0     0     0     0     0     0     0     0     0     0     0     0
CINDY       0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   692     0   429
DIANE       0     0     0     0     0   354     0     0     0     0     0     0     0     0   449     0     0     0   408     0
EGOR        0     0     0     0     0     0     0   367     0     0     0   388     0     0     0     0     0     0     0     0
FRAN        0     0     0   454     0     0     0     0     0     0     0     0     0     0     0     0     0     0   700     0   3.
GREG      323   454     0     0     0     0     0     0     0     0     0     0     0     0     0   311   371     0     0     0
HENRY     324     0     0     0   363     0     0     0     0     0     0   359     0     0     0     0     0     0     0     0
IONA        0     0     0     0     0     0     0     0     0     0     0     0   445     0     0     0     0     0     0     0
JASON       0     0     0     0     0     0     0     0     0     0     0     0     0   371     0     0     0     0     0     0
KELLI       0     0     0     0     0     0     0     0   315     0     0     0   320     0   441     0     0     0     0     0
LARRY       0     0     0     0   396     0     0   375     0     0     0     0     0     0     0     0     0     0     0     0
MARY        0     0     0     0     0     0     0     0   443     0     0     0     0     0     0     0     0     0     0     0
NICK        0     0     0     0     0     0     0     0     0   424     0     0     0     0     0     0     0     0     0     0   3
OLIVE       0     0     0   331     0     0     0     0     0     0   457     0     0     0     0     0     0     0   305     0
PETER     464     0     0     0     0     0   323     0     0     0     0     0     0     0     0     0   312     0     0     0
QUINT     331   397     0     0     0     0   506     0     0     0     0     0     0     0     0   413     0     0     0   317
RITA        0     0   691     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0   434
SANDY       0     0     0   350     0   463     0     0     0     0     0     0   319     0   354     0     0     0     0     0
TERRY       0     0   524     0     0     0     0     0     0     0     0     0     0     0     0     0   368   531     0     0
UXTON       0     0     0     0     0     0     0     0     0     0     0   357     0   394     0     0     0     0   300     0
```

```
DAY    ----------------------------------------------- (SCALED) KID TO TEACHER-------------------------------------------------- 13-NOV-76    08:5
        ANDY  BEN  CINDY DIANE EGOR  FRAN  GREG HENRY IONA JASON KELLI LARRY MARY  NICK OLIVE PETER QUINT RITA SANDY TERRY UATI
   0    400   400   400   400   400   400   400   400   400   400   400   400   400   400   400   400   400   400   400   400   40
   4    408   434   405   391   386   341   361   366   405   363   351   408   424   357   296   447   350   490   326   459   34
   8    413   446   316   381   361   258   309   304   307   309   274   415   442   273   242   500   350   587   227   523   24
  12    404   551   310   330   344   195   257   306   293   260   246   308   349   166   225   337   383   473   155   567   19
  16    393   638   328   261   326   153   213   286   237   218   238   262   318    95   188   292   413   511    99   595   17
  20    295   705   344   233   306    94   183   213   182   188   194   264   318    72   164   314   437   563    59   612   14
  24    278   554   266   178   246    43   163   202   186   169   155   283   341    47   164   351   460   616    2    466    9
  28    281   585   267   175   201    20   135   224   206   150   158   304   360     6   173   264   346   652    -1   453    6
  32    290   640   285   189   200    -5    96   192   234   120   175   243   278   -35   165   216   297   685   -24   478    5
  36    234   693   305   160   208   -13    81   122   261    79   140   230   275   -37   105   231   306   633   -37   504    7
  40    224   580   255   145   216    -8    82    95   275     9   134   255   291   -28    62   275   333   482   -31   528    4
  44    239   503   181   162   223    -6    78   114   180   -40   142   284   307    -7    55   327   366   481   -25   412
  48    228   535   168   179   146   -14    45   146   119   -45   123   307   213     0    76   262   240   520   -65   423   -
  52    191   595   190   186    83   -67    -2   123   100   -35    73   329   198    11    69   294   173   567  -117   404   -
  56    209   655   148   204    60   -93   -18   154    80   -22    40   276   188    22    75   349   173   478  -143   294   -
  60    236   706   138   161    75   -98   -11   196   100   -25    43   211   158    45   106   397   207   410  -141   268   -1
  64    264   663   168   148 [illegible] -93  11   150   128   -27    46   206   181    44   135   432   251   442  -113   304   -
  68    284   472   213   158    50   -68    41   118   115   -20    24   171   215    34   162   352   256   497   -72   362   23
  72    200   454   255   149    29   -64    66   125   128   -17    27   141   248    32   145   345   234   555  -104   417   27
  76    187   499   213    97    50   -56    62   153   163     0    60   159   283    29    90   373   268   414  -153   331   34
  80    208   560   224    92    81   -45    27   198   197    -5    70   194   276    35    44   403   307   344  -171   355   15
  84    241   620   258    90    69   -30    -5   240   230   -12    82   167   207    12    33   427   345   371  -137   407   1
  88    185   461   260    53    56   -30    -3   198   203     3   114   180   202    -6    52   323   378   433  -123   454   17
  92    166   398   198    52    70   -76    19   205   192    11   105   213   222    -4    67   281   295   497  -121   496   37
  96    185   431   206    70    93   -89    48   228   212    13   114   245   245   -26    50   293   304   544  -123   527   52
 100    215   500   236    96   122   -77    40   252   239    16    89   267   273   -48    57   280   338   444   -93   425   34
 104    248   574   278    88   155   -50    17   166   228    10    52   198   264   -48    94   219   371   447   -77   412   33
 108    280   481   242    99   187   -42    14   139   199    28    55   141   182   -44   134   234   313   490   -79   442   38
 112    305   504   173   124   122   -38    27   152   214    49    85   135   154   -30   142   271   210   387   -74   475   66
 116    232   579   173   157    82   -21    40   177   184    73    56   153   160   -13    94   275   184   411   -59   504   43
 120    240   650   212   160    88    -8    39   218   149    59    25   155   195     6    64   248   210   461   -55   526   29
 124    266   568   258   139   123    13    53   197   161    56    26   146   233    26    66   281   258   506   -60   542   25
 128    294   537   208   152   161     8    74   191   190    70    53   175   153    36    91   318   202   359   -52   381   31
 132    240   588   234   162   125    15    59   223   175    75    98   140   109    23   103   356   194   350   -58   353   54
 136    239   643   276   142   141    33    27   224   166    44    84   153   114    20    68   302   170   391   -67   372   73
 140    267   614   318   131   168    38    24   202   194     4   102   194   132    28    28   250   155   444   -63   359   93
 144    297   545   349   145   195    38    41   229   236    -1   133   230   125    34    19   259   189   372   -49   255   84
 148    318   585   373   164   158    66    68   255   178    25   170   247   135    33    36   291   206   386   -37   239   81
 152    298   641   289   142   155    89    99   221   142    17   148   195   167    24    42   325   199   441   -48   278  105
 156    204   561   293    96   178    58   129   161   143   -17   149   208   203    36    53   314   245   495   -59   333   81
 160    177   541   320    98   199    17    89   150   169   -17   167   231   240    63    69   237   301   539   -60   273   88
 164    192   590   348   125   167     1    79   177   211   -13   170   215   197    48   111   239   356   385   -53   311  103
 168    223   639   374   159   128     7    94   213   209    23   143   181   213    32    86   272   351   377   -44   370  130
 172    187   526   301   121   126    18   122   196   186    15   160   198   239    41    91   311   310   417   -41   425  142
 176    194   528   292    70   129    19   102   195   203    -2   131   205   269    66   122   344   336   472   -36   352  128
 180    228   582   316    48    99    39   105   230   229     0   155   200   216    45   156   332   376   522   -22   359  106
 184    183   633   343    56   102    63   133   270   254    28   155   133   213    41   190   288   268   499   -16   400  107
 188    201   597   366    64   122    70   158   301   280    34   131   113   245    61   147   308   269   435   -51   443  126
```

6.0 Summary and Conclusions

6.1 Summary

This report has dealt with a wide range of topics. As discussed in Chapter 1, the project began with a recognition by this investigator that some fairly serious problems were endemic in the perceived utility of educational research. These problems affected research producers and research users alike, and did so in a variety of ways. The basic nature of the problem appeared to be a failure of some sort in the inferential link between data and theory or model. It was discovered that the alternative paradigm of statistical inference known as the Bayesian viewpoint showed clear signs of being able to alleviate several aspects of this generic inference and credibility problem. Thus, the present project was undertaken to investigate the applicability of this new paradigm to one important area of education research, the evaluation of compensatory education programs.

Chapter 2 of the Report reviews in some detail the key features of the Bayesian paradigm of statistical inference. The fundamental distinguishing feature of this paradigm is not a statistical idea at all, but rather involves one's conceptualization of what knowledge really means in a context where chance events occur. We developed at some length the idea that in such circumstances, our knowledge itself becomes "probabilistic", and is best represented by a set of "probability" values, or a "probability" curve. However, partly to emphasize the fact that these curves are different in nature from the ideas usually implied by that word, we labeled them as "likelihoods", and developed the concept of a standardized conditional likelihood (SCL). The way in which these concepts are involved in Bayesian statistical inference also was described, as was the Achilles heel of Bayesian inference -- how to formally represent ignorance as one point on a range of knowledge.

The material of Chapter 3 was much more down-to-earth in tone. As a result of our searching for available computer packages which would be convenient to use to carry out the mechanics of Bayesian statistical inference, we became convinced that no truly suitable packages had been developed. Two packages, CADA and MANECON, each were of some value, but were not fully adequate for our purposes. Consequently, several computer programs were developed as part of this project. These are described in Chapter 3. The programs include two which carry out evaluation of important distributions often used in Bayesian analyses: (1) TBF, which evaluates the Student's t and the Behrens-Fisher distributions; and (2) ICD, which evaluates the inverse-chi distribution. All of these distributions arise naturally in connection with Normal distribution data-generating models. Additional programs (INTEG and BFTEST) were developed to investigate techniques of numerical integration and to employ such techniques in validating an aspect of the TBF algorithm.

In Chapter 4 of the Report an actual application of Bayesian inference, using the materials discussed in Chapter 3, is presented. This application is a reanalysis of data on achievement outcomes from the O.E. Incentives Project. The reanalysis indicates that Bayesian techniques tend to provide considerably clearer inferences when the sample size is small. This does not mean that the inferences are definitive, but simply that they are more properly consistent with the actual data and a "commonsense" interpretation thereof. The reanalyses also showed that the interpretation of such studies is highly sensitive to the particular beliefs of the person making the interpretation. Some conceptual issues as to the proper model specification, especially that of the level of aggregation, also were discussed in Chapter 4.

The succeeding chapter, Chapter 5, described how one discovery of the project was the need for formal, explicit specification of a complex, overall model for the data-generating process. In particular, a distinction was drawn between the "proximate" model and the more "global" context within which it operates. To accomplish such specification, a computer simulation approach seems necessary. As one step toward developing that sort of approach, we developed a model and a corresponding program we called the Classroom Dynamics model. This model and the program are presented in some detail in Chapter 5 and its Appendix 5A. Finally, there are some important and fundamental issues raised by this use of simulation as to the most suitable strategies for inference, and what our aspirations or expectations about such inference should be. The last section of Chapter 5 offers some ideas on these topics.

6.2 Conclusions

As with so many other research projects, this one draws as a major conclusion the idea that further investigation is needed. But that oversimplifies the matter.

One basic conclusion of our work is that the Bayesian inference approach is currently quite feasible but not currently ripe for use. That is, the approach depends on the availability of cheap, interactive computing facilities, and of suitable data analyses packages. We conclude that the hardware is here today, and is constantly improving. The programs, however, are in a primitive stage of development. Our work has advanced that development a few steps but much more is needed. Such work will require the involvement of specialists in areas such as numerical analyses methods as well as those of educators and statisticians.

A second basic conclusion is that the Bayesian inference analysis does serve to reveal facets of the data neglected by conventional approaches. To say it perhaps too simply, the Bayesian approach, by reminding us of the probability aspect of all our inference, helps us to use more fully and correctly data which is diagnostically weak.

The third and last basic conclusion is an unanticipated one. It involves the formalization of our process models for the phenomena under study. The Bayesian paradigm causes us to take such models quite seriously as a potentially sensitive link in the chain of inference. When we do so, it becomes apparent that, though computer simulation is a major aid in simply representing such models faithfully and consistently, there is need for some radical change in our views about the nature and limits of scientific inference with respect to truly complex, adaptive, stochastic systems such as those of the classroom or school.

All of these conclusions are sobering ones, but not disappointing. They lead us to feel that the communication-credibility problem that prompted this present project is now itself much better understood. More to the point, the work has also enabled us to identify several specific points at which future investment of resources is needed, and so to plan more intelligently for the improvement of educational administration at all levels.

REFERENCES

Abramowitz, M. and I. A Stegun, editors.
1968 *Handbook of Mathematical Functions*. Washington, D.C.: National Bureau of Standards, Applied Mathematics Series, 55; U. S. Government Printing Office.

Aspin, Alica A.
1949 "Tables for Use in Comparisons Whose Accuracy Involves Two Variances, Separately Estimated". *Biometrika* 36: 290-293.

Box, G. E. P. and G. C. Tiao.
1973 *Bayesian Inference in Statistical Analysis*. Reading, Massachusetts: Addison-Wesley.

Braithwaite, R. B.
1953 *Scientific Explanation*. Cambridge. Cambridge University Press.

Brownlee, Kenneth A.
1965 *Statistical Theory and Methodology in Science and Engineering*. Second Edition. New York: Wiley.

Carnap, R.
1950 *Logical Foundations of Probability*. Chicago. University of Chicago Press.

Carroll, J. B.
1963 "A Model of School Learning". *Teachers College Record* 64: 723-733.

Coleman, James
1973 *The Mathematics of Collective Action*. Chicago. Aldine.

Coleman, James S. et al.
1966 *Equality of Educational Opportunity*. Washington, D.C., U. S. Government Printing Office.

Coleman, James
1964 *Introduction to Mathematical Sociology*. New York. Free Press of Glencoe.

Collet, Lee
1974 *FEHR-Practicum*. Final Report to U. S. Office of Education, Ann Arbor, Michigan. The University of Michigan.

Cooley, William W., and Paul R. Lohnes.
1976 *Evaluation Research in Education*. New York. Irvington-Halsted Press.

Cyert, Richard M., and James G. March
1963 *A Behavioral Theory of the Firm*. Englewood Cliffs, New Jersey. Prentice-Hall.

Fennessey, James

1974 "Understanding 'Fan-Spread' in Achievement Measures". Report #168. Center for Social Organization of Schools. The Johns Hopkins University.

1972 "Some Problems and Possibilities in Policy-Related Social Research". *Social Science Research* 1: 359-383.

1970 Determinants of Judgments among Elementary School Teachers: Theoretical, Methodological, and Empirical Explorations. Unpublished Ph.D. dissertation. The Johns Hopkins University. Baltimore, Maryland.

Fisher, R. A.

1925 "Expansion of Student's Integral in Powers of n(-1)". *Metron* 5: 109-112.

Fisher, R. A. and M. J. R. Healy

1956 "New Tables of Behrens' Test of Significance". *Journal of the Royal Statistical Society*, Series B 18: 212-216.

Fisher, R. A. and F. Yates

1957 *Tables for Statisticians*, Fifth Edition. New York: Hafner.

Fisher, R. A.

1973 *Statistical Methods and Scientific Inference*, 3rd Edition. New York: Hafner.

Forrester, Jay

1968 *Principles of Systems*. Cambridge, Massachusetts. Wright-Allen Press.

Gallaher, L. J.

1973 "A Multidimentsional Monte Carlo Quadrature with Adaptive Stratified Sampling". Algorithm 440. *Communications of the Association for Computing Machinery* 16: 49-50.

Hacking, Ian

1975 *The Emergence of Probability*. Cambridge. Cambridge University Press.

Hampton, J. M., P. G. Moore and H. Thomas

1973 "Subjective Probability and its Measurement". *Journal of the Royal Statistical Society*, Series A 136: 21-42.

Heyns, Barbara Lee.

1971 Curriculum Assignment and Tracking Policies in Forty-Eight Urban Public High Schools. Unpublished doctoral dissertation. The University of Chicago. Chicago, Illinois.

Hill, G. A

1970 "Student's $\underline{t}$ Distribution". *Communications of the Association for Computing Machinery* 13: No. 10, pp. 617-619.

Hill, I. D., and M. C. Pike
1967 "Chi-Squared Integral". Algorithm 299. *Communications of the Association for Computing Machinery* 10: 243-245.

Hogarth, Robin M.
1975 "Cognitive Processes and the Assessment of Subjective Probability Distributions". *Journal of the American Statistical Association* 70: 271-291.

Isaacs, Gerald L. and Melvin R. Novick
No Date "The Bayesian Computer-Assisted Data Analysis (CADA) Monitor". The University of Iowa.

Isaacs, Gerald L., et al.
1974 *Tables for Bayesian Statisticians*. Iowa City. The University of Iowa.

Jeffreys, Harold
1961 *Theory of Probability, Third Edition*. Oxford. Oxford University Press.

Johnson, N. L. and S. Kotz
1970 *Continuous Univariate Distributions, Volume 1 and Volume 2. Boston: Houghton Mifflin (distributed by Wiley)*.

Keynes, John M
1921 *A Treatise on Probability*. London. MacMillan & Co., Ltd.

Kuhn, Thomas M.
1970 *The Structure [illegible] Scientific Revolutions, Second Edition, Enlarged*. Chicago. The University of Chicago Press.

Land, Kenneth
1971 "Formal Theory". in Herbert L. Costner, editor. *Sociological Methodology 1971*. San Francisco. Jossey-Bass.

Levine, David A.
1969 "Student's *t* Distribution". *Communications of the Association for Computing Machinery* 12: No. 1, pp. 37-38.

Lindley, Dennis V.
1965 *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge. Cambridge University Press.

Majumder, K. L. and G. P. Bhattacharjee
1973 "The Incomplete Beta Integral". *Applied Statistics* 22: 409-411.

March, James G., and Herbert A. Simon
1958 *Organizations*. New York. Wiley.

Mischel, Walter
1973 "Toward a Cognitive Social Learning Reconceptualization of Personality". *Psychological Review* 80: 252-283.

Morris, John
1968 "T-Test Probabilities". Communications of the Association for Computing Machinery 11: No. 2, pp. 113-116.

Novick, M. R. and P. H. Jackson
1974 Statistical Methods for Educational and Psychological Research. New York: McGraw-Hill.

Patil, V. K.
1965 "Approximations to the Behrens-Fisher Distributions". Biometrika 52: 267-271.

1964a "Difficulties Involved in Computing Behrens-Fisher Densities, Cumulative Probabilities, and Percentage Points from First Principles". Journal of the Indian Statistical Association 2: 109-118.

1964b "The Behren-Fisher Problem and Its Bayesian Solution". Journal of the Indian Statistical Association 2: 21-31.

1963 The Behrens-Fisher Distributions. Unpublished Ph.D dissertation. Department of Mathematics, The University of Michigan, microfilm #63-6937.

Pearson, E. S. and H. O. Hartley
1966 Biometrika Tables for Statisticians, Volume I. Cambridge: Cambridge University Press.

Planar Corporation
1972 Incentives in Education Project Impact Evaluation Report. Washington, D. C. Planar Corporation.

Platt, John R.
1964 "Strong Inference". Science: 146, 347-352 (14 October)

Raiffa, Howard
1968 Decision Analysis. Reading, Massachusetts. Addison-Wesley.

Raiffa, Howard, and Robert Schlaifer
1961 Applied Statistical Decision Theory. Cambridge, Massachusetts. Graduate School of Business Administration. Harvard University.

Renyi, Alfred
1955 "On a New Axiomatic Theory of Probability". Acta Mathematical Academy of Sciences of Hungary 6: 285-335.

Salmon, Wesley C.
1971 Statistical Explanation and Statistical Relevance. Pittsburgh. University of Pittsburgh Press.

Savage, Leonard J.
1964 "The Foundations of Statistics Reconsidered". in Henry E. Kyburg and Howard E. Smokler, editors, Studies in Subjective Probability. New York. Wiley.

Schlaifer, Robert
1971 *Computer Programs for Elementary Decision Analysis*. Cambridge, Massachusetts: Harvard University Press.

Schmitt, Samuel
1969 *Measuring Uncertainty*. Reading, Massachusetts. Addison-Wesley.

Snedecor, George W. and William G. Cochran
1967 *Statistical Methods, Sixth Edition*. Ames, Iowa: Iowa State University Press.

Soper, H. E.
1921 "The Numerical Evaluation of the Incomplete Beta-Function." in *Tract for Computers, No. 7*. Cambridge: University Press.

Spilerman, Seymour
1971 "The Causes of Racial Distrubances: Tests of an Explanation". *American Sociological Review* 36: 427-442.

Tanburn, Elizabeth.
1938 "Note on an Approximation Used by B.L. Welch". *Biometrika* 29: 361-362.

Von Neumann, John and Oskar Morgenstern
1947 *The Theory of Games and Economic Behavior, Second Edition*. Princeton. Princeton University Press.

Welch, B. L.
1949 "Further Note on Mrs. Aspin's Tables and on Certain Approximations to the Tabled Function". *Biometrika* 36: 293-296.

1947 "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved". *Biometrika* 34: 28-35.

1938 "The Significance of the Difference Between Two Means when the Population Variances are Unequal". *Biometrika* 29: 350-361.

Westinghouse Learning Corporation/Ohio University
1969 *The Impact of Head Start*.

Wiley, David E., and Annegret Harnischfeger
1975 "Teaching-Learning Process in Elementary Schools: A Synoptic View". Report #9, Studies of Educative Processes.

Winer, B. J.
1971 *Statistical Principles in Experimental Design, Second Edition*. New York. McGraw-Hill.

Winkler, Robert L.
1967 "The Assessment of Prior Distributions in Bayesian Analysis". *Journal of the American Statistical Association* 62: 776-800.